

Japan and its women

Cultural obstacles and potential damage to one's career present major challenges to women wanting to pursue science in Japan. Some changes have occurred, but too few, and too slowly.

Why do Japanese women scientists have such a hard time of it? In their country, it seems, one is expected to wait for change to happen, not to force it, and women are supposed to *gaman*, or endure. Improvements in such matters as equal opportunity are thought of as gifts to be bestowed by a leader.

Some say that the Japanese prefer harmony to conflict and mediation to open courtrooms, and so do not actively pursue change. It is debatable, and hotly debated in Japan, whether this reflects a widely held cultural value or whether such norms are the result of circumstances that make contention difficult, such as a prohibitively expensive and inaccessible legal system. Is this culture of 'harmony' merely a tool used by government leaders, management, or whoever else might find it handy in dealing with those who make a fuss?

Certainly, it is not easy to successfully raise a complaint in Japan. If a woman feeling unjustly slighted in a job or grant competition complains, she is considered to be putting her own career interests ahead of the group's harmony. She might even be told that she should make room for males, and should concentrate on her home life. Alternatively, a woman might lose grants or fail to obtain a position because a decision-maker assumes her research will suffer because she will spend too much time with children she may not even have. Even those who support women might not be understanding. "Concentrate more on research, less on career objectives — do good research, and that will carry you through." The assumptions underlying this train of thought are that women have equal opportunities to do research and that their achievements will be evaluated objectively. Clearly, this has not been the case, as the testaments of many women show (see page 404).

Older Japanese men tend to acknowledge under-representation by women (a quick look at the statistics would convince anyone), but generally they think of it as a problem of choice — most women, as

they see it, choose to have children and thus their careers do not progress. They fail to see that a woman's career could be affected by their own attitudes — attitudes that might colour evaluations on employment or grants awards and that produce offhand remarks about "a woman's place" made to aspiring women postdocs or graduate students. Younger men in Japan tend to be more conscious of the problem, and more sympathetic. This is especially true of those who may have watched their wife struggle in this system. They are more comfortable with women as colleagues, and some take on child-care responsibilities to make their wife's career possible.

There are some promising developments in Japan's scientific societies. A few women have reached high rank, and several societies have women's sub-groups. Women in Physiology in Japan, for example, releases a newsletter that conducts and reports on surveys as well as discussing issues relevant to women's careers (<http://web.kanazawa-u.ac.jp/~med2/05/WPJ-MENU.html>; only in Japanese). Drawing attention to the problem will continue to create pressure for change. An increasing number of societies are setting up child-care facilities at conferences to allow researcher-mothers to attend.

Some universities and institutions, such as the Institute of Physical and Chemical Research (RIKEN), seem to have created a comfortable place for women to work, but female representation more generally is abysmal. If the government wants to stand by its word and increase the number of women in research, it must go beyond current proposed concessions, such as allowing a woman to apply for a grant using her maiden name. Government agencies, starting with the Ministry of Education, Culture, Sports, Science and Technology, need to ensure open and fair employment decisions and grant awards.

Gradual changes in favour of women will no doubt continue. But how many promising careers will be dashed, and what toll will this take on Japanese science in the meantime? ■

Collider's moment of truth

A German proposal for a large linear accelerator is welcome. A global approach would be even more so.

With this week's proposal from DESY for a 500-giga-electron volt linear accelerator (see page 397), the time has arrived for high-energy physicists to start thinking seriously about how, when and where such an accelerator should be built.

Three principle efforts to design it are under way: DESY's superconducting TESLA, and the more conventional Next Linear Collider, being developed in the United States, and Japan Linear Collider. The question of choosing the right design should be made independently of the deeply political decision of where to locate it. Sadly, in the world of big science, such distinctions are rarely clear-cut.

For US physicists, still smarting from the cancellation of the Superconducting Super Collider (SSC) in Texas in 1993, decision time is perhaps arriving a couple of years too soon. The expenditure of several billion dollars on the SSC before its cancellation has left an unfortunate gap between the likely cost of the Next Linear Collider — estimated at \$7.9 billion in 1999 — and the amount the US govern-

ment will realistically support. Meanwhile, Japanese physicists, who deservedly have held high hopes of hosting a large linear collider, face a changing political climate in which funding may be increasingly difficult to obtain.

In Europe, where official cost estimates for a linear collider may conceal an actual price (of the sort required under US law, which includes the costs of labour and operations) similar to that ascribed to the Next Linear Collider, it is unlikely that sufficient support can be gathered from the German government or through a European collaboration, not least given the continuing demands of CERN.

The way forward is to develop an effort that is genuinely global from the outset. There are at least two prerequisites for successful collaboration, neither of which yet exists: robust political support at the highest levels, and an acceptance by Europeans that, with CERN's Large Hadron Collider, they have had their turn. The latter could help high-energy physicists worldwide acquire the former. ■





Sky high
Spiralling costs of
space station put
research in doubt
p399



War of words
Eruption books
reignite volcano
researchers' feud
p400



Cool on warming
US administration
calls climate change
science 'incomplete'
p401



Crossed wires
The Vatican's radio
transmitters fall
foul of regulations
p402

German lab unveils plan to build physicists' next particle collider

Alison Abbott and Josette Chen

The next era of high-energy physics will come into sharp focus this weekend when an international team of physicists unveils a proposal for a major linear collider.

The team, based at the DESY particle physics laboratory near Hamburg, will propose the construction of a machine called TESLA — the Tera electron volt Energy Superconducting Linear Accelerator — at an estimated cost of about US\$3 billion. The machine would be a linear electron-positron collider based on superconducting resonators (also known as 'cavities'). It would allow particles to collide at energies of between 500 and 800 gigaelectron volts (GeV).

The announcement will attract keen global interest because both the United States and Japan believe they should host the next major collider project. Physicists from both countries have been jointly developing a proposal for a linear collider that would meet similar energy goals to those of TESLA using conventional, rather than superconducting, resonators to accelerate the electrons. The project is known in the United States as the Next Linear Collider (NLC) and



Linear thoughts: DESY staff assemble prototype accelerator parts for the proposed TESLA collider.

in Japan as the Japan Linear Collider (JLC).

The various teams, all of them international, are downplaying the competitive aspect, but each community of physicists would obviously like the next-generation machine to be built on its own territory. Both the United States and Japan are making substantial contributions to another major project, the Large Hadron Collider (LHC) that is being built at CERN, the European Laboratory for Particle Physics, near Geneva. Construction of TESLA in Germany, with the LHC also in Europe, could mark the virtual

eclipse of particle physics in the United States and Japan for a generation.

A linear collider would help physicists to discover what lies beyond the standard model, the theory that describes all elementary particles and the corresponding forces between them. Many aspects of the theory have been fully validated. But one exception is the Higgs boson, the heavy particle that physicists hope to find using either the renovated Tevatron at Fermilab in Illinois (see *Nature* **409**, 754–755; 2001) or the LHC.

Both these accelerator rings, which create collisions of heavy particles such as protons, may detect evidence for supersymmetry, a theory that would augment the standard model by finding corresponding 'sparticles' for each known particle. But proton colliders produce messier and more complex collisions than electron-positron colliders, which have long been used to complement them in particle physics investigations.

"The LHC will provide one view beyond the standard model and the electron-positron collider will provide another," says Burt Richter, former director of the Stanford Linear Accelerator Center (SLAC) in California, which has led the US arm of the NLC design.

Both TESLA and the NLC would achieve high enough energies — in the 500 GeV to 1 TeV ($\times 10^{12}$ eV) range — to complement the LHC. But the NLC design uses conventional technologies developed at SLAC and at the Japanese laboratory KEK, which are based on copper resonators. NLC advocates say

No Wellcome money for Celera

David Adam, London

The Wellcome Trust, the world's largest medical charity, has banned its grant recipients from using its funding to subscribe to the human genome sequence prepared by Celera Genomics. Instead it wants them to use the public human genome sequence, which it partially funded and helped to prepare.

A Wellcome policy statement says that 'no trust funds, including contingency monies, may be used for the purpose of accessing Celera subscription services'. Trust-funded researchers accessing Celera's sequence must do so with money 'from sources wholly independent of the trust.'

A spokesperson for Celera said that

Wellcome Trust researchers would still be able to access the company's data: "We have a free site and so presumably they will have access that way."

"There is no evidence that Celera's database offers any scientific advantage," says Mike Dexter, director of the trust. "This is not a ban: we are just trying to get the best value for trust money."

The trust uses revenues from its £15 billion (US\$21 billion) asset base to fund biomedical research in UK universities and, to a lesser extent, in the developing world. It distributes about £600 million in grants every year, including funding for the Sanger Centre in Cambridge, the UK arm of the Human Genome Project.

▶ that this will allow the machine to operate at higher frequencies, yielding faster acceleration and a smaller, and therefore potentially cheaper, facility than TESLA.

Another advantage of the NLC design is the technology's established track record. "We already have one at 50 GeV, so going to 500 GeV is only a factor of 10, and we have lots of experience," says Richter.

But the NLC concept received a blow in the past year, when researchers found physical deterioration in the copper cells after 1,000 hours of operation. "We should know by the summer if new designs will overcome these problems," says Steve Holmes, associate director at Fermilab.

More ominously, an initial 1999 cost estimate for the NLC came in at a politically unrealistic \$7.9 billion. "We need this to be reduced by 25%, preferably by 50%," says Peter Rosen, head of high-energy physics at the US Department of Energy. He says a potential reduction of 30% has already been achieved.

TESLA's use of superconducting resonators should mean a higher beam intensity and lower operating costs, its advocates say. Compared with the NLC, the radiofrequency devices that excite superconducting resonators are simpler to build. But the largest functioning accelerator using superconducting technology, the Thomas Jefferson National Accelerator Facility in Virginia, operates at only 1 GeV.

Albrecht Wagner, DESY's director, believes that the price estimate his team will publish on 23 March will be considered reasonable. But unlike the US estimate, it will exclude labour and operational costs, which could double the actual price.

The TESLA proposal will be assessed by the Wissenschaftsrat, Germany's science council, which will decide whether to recommend it to the government. German research minister Edelgard Buhlman has made positive noises about TESLA, but has said the project will go ahead only with the support of the international particle physics community — and financial support from abroad. But Wagner contends that construction could start within two years.

Political factors are likely to settle the choice of both the site and the design. The US community plans a major gathering of particle physicists in Snowmass, Colorado, this July to look at future priorities. A sub-panel of the High Energy Physics Advisory Panel is meanwhile mapping out a 20-year plan for the discipline.

With substantial funding allocated for building the LHC, Europe is not flush with money for a new collider either, and DESY is trying to promote TESLA as a multi-disciplinary facility whose associated free-electron laser will also be used by materials scientists, chemists and biologists. ■

More culls planned as Britain wrestles with foot-and-mouth

David Adam, London

Up to a million healthy farm animals are to be slaughtered in a fresh attempt to break the stranglehold of foot-and-mouth disease on the British countryside, the government announced last week.

All sheep and pigs within three-kilometre exclusion zones around areas of high infection will be culled. Government vets hope that the mass slaughter will create 'fire-walls' that the infection cannot cross.

Meanwhile, with more than 325 cases of the disease now reported, officials have admitted that attempts to contain the virus among animal herds that were infected before movement restrictions were introduced have failed.

"Some of the new cases we're seeing are down to contiguous infection, things like airborne spread, vehicle movement and people movement," says a spokesman for the Ministry of Agriculture, Fisheries and Food (MAFF).

But he denies that the outbreak is out of control, pointing out that new cases can still be traced back to animals infected during the original outbreak. "We're still saying that [the epidemic] is under control to a certain extent," he says.

But some experts disagree. "The crucial measure is if the transmission potential is self-sustaining," says Roy Anderson, a leading epidemiologist at Imperial College, London. Because each primary infection on a farm is leading to more than one secondary infection, he says, the situation "is quite clearly not under control".

Anderson agrees that the mass slaughter around infection hot-spots is "absolutely the right thing to do". But he is worried about whether other, more isolated cases can be effectively contained.

The epidemic, which broke out last month (see *Nature* **410**, 4; 2001), is the first major outbreak in Britain since 1967. One reason why it has proven to be more difficult to control than the government had hoped is that some infected animals were sold in private deals at cattle markets, so the sales were never officially recorded and the animals could not be traced until their symptoms were observed.

Although the initial stages of any epidemic are the hardest to predict, epidemiologists remain optimistic about the possible control of the disease in mainland Europe. The first case of foot-and-mouth was confirmed in France last week. But France traced and slaughtered all animals that had been imported from Britain in the weeks preceding the initial outbreak, and placed herds



that may have been in contact with them under quarantine and close surveillance. The infected animal was found in one of these herds.

"The French have reacted very quickly and very well," Anderson says. "An epidemic is a non-linear process, therefore action taken in the early stages prevents far more cases than action taken later."

Mark Savay at the National Veterinary Centre in Paris adds that restrictions on animal meetings and transportation were introduced in France within days of the first UK case. "We received clear instructions from the European Commission to do that," he says, adding that other European Union countries would have received orders to do the same.

Controls on meat, livestock and milk imports from Britain, as well as animal movement restrictions, have been tightened in most other European countries. Countries outside Europe, including Japan and the United States, are also acting to keep the virus out, banning imports from Europe and introducing strict checks on travellers.

Vets and epidemiologists agree that modern patterns of animal movement and trade have increased the spread of the infection. It has already spread further in Britain than the 1967–1968 outbreak did in six months. "The number of animals moved and the number of journeys is not that different today but the distances involved are greater," says Chris Bostock, director of the government-funded Institute for Animal Health. ■

▶ <http://www.maff.gov.uk/animalh/diseases/fmd/default.htm>

Space-station cuts leave research in lurch

Tony Reichhardt, Washington

Scientists who hope to work on the International Space Station fear that proposed budget cuts will severely impair the orbiting laboratory's research potential.

The researchers say that a new attempt to cut costs represents the biggest setback for the project since its last major redesign in 1993. The cost controls were introduced by the incoming Bush administration when it found that the station's price tag had soared to \$4 billion more than its planned level.

The White House has ordered NASA to find \$4 billion in savings over the next five years to offset the projected cost overruns. The savings are almost certain to reduce and delay research on the station.

According to one advisory group, if the impact on the station's capability is too drastic, "there will no longer be adequate justification for continued support of [the station] by the scientific community at large". The warning came from the scientific panel that oversees the Space Station Biological Research Project, a suite of lab facilities being developed at NASA's Ames Research Center near San Francisco, which is due to be installed on the station in stages over the next several years.

The working group's chairman, Martin Fettman of Colorado State University, wrote in a letter to NASA's chief of human space-flight on 9 March that the group "is prepared to resign, effective immediately" if the research programme is cut back too far. "We believe the entire life-sciences community could turn its support away from [the station], and in fact become active campaigners against [it]" if the programme "continues to divert resources from science to solve construction problems".

The leverage exercised by such a warning may be limited by the fact that many scientists have opposed the station since its inception in 1984 (see *Nature* **307**, 1–2; 1984). But the latest rumblings come from individuals who have stuck with the project and supported some of NASA's claims for the viability of its research.

NASA plans to save half of the required \$4 billion by cancelling three key elements of the space station: a crew habitation module, a US-built rescue vehicle, and a propulsion module. Scrapping the first two means the station could house only three astronauts at a time instead of the planned six, unless some other rescue vehicle is attached to the station. Crew time is vital for conducting research, so this would dramatically affect the number of experiments that could be done, particularly in the biomedical sciences.

To make up the other \$2 billion, station managers are looking at several options, including cancelling or delaying some of



Over budget: the International Space Station must make savings of \$4 billion after cost overruns.

the large 'facility-class' research hardware planned for the station, or shifting the experiments to smaller, modular racks of the kind flown on past space-shuttle flights. But researchers say that either action would diminish the station's value as a laboratory.

Speaking to the National Academy of Sciences' Space Studies Board last week, Steve Isakowitz of the White House Office of Management and Budget said the cost overruns meant that the Bush administration effectively is committing only to a "core complete" version of the station, rather than the "assembly complete" version that was planned.

When assembly flight 10-A is completed in late 2003, NASA will have satisfied three key criteria, Isakowitz said. The station will be permanently occupied, the agency will have fulfilled its obligation to provide an attachment point for international modules, and the station will be better than any previous research facility in orbit. But any

enhancements to the core design will have to come out of NASA's existing budget.

Some of the options under discussion have left life scientists in a near panic. One scenario entails scrapping a centrifuge considered essential for doing controlled studies of gravity's effect on animals and plants, along with much of the biological research equipment scheduled to be installed in the module that houses it. The module would instead be converted to living quarters for the crew, raising it back up to six.

Kathie Olsen, NASA's chief scientist, told the Space Studies Board that some of the options under discussion "would have a significant impact on the types of science that we'll be able to do", and may force the agency to eliminate entire fields of research from the station. But NASA officials advise patience, and hope that the station will eventually retain its capacity for a full crew and research programme. ■

Israel plans new particle accelerator

Haim Watzman, Jerusalem

A new particle accelerator will replace the 40-year-old research reactor at Israel's Soreq Nuclear Research Center by 2006, the government has announced.

The \$24 million facility will be used primarily as a source of neutrons, including cold neutrons, for materials scientists, structural biologists and other researchers, and for the production of isotopes.

Construction of the Soreq Applied Research Accelerator Facility (SARAF) is one of the first projects initiated by Telem, the Israeli National Forum of Research and Development Infrastructure.

Although still under design, SARAF is expected to consist primarily of a 50–70 mega-electron volt particle accelerator that will be able to generate proton beams with an intensity of 1–2 mA. The beam will be directed at a target to produce neutrons, whereas the old reactor produced them directly. The project's management will decide shortly whether the beam will come from a ring or a linear accelerator.

Zvi Kaplan, director of the Soreq centre, hopes to attract visiting research groups from the United States, Europe and Japan for pilot experiments to help Israeli researchers prepare the accelerator for use. ■

Conflicting volcano tales erupt in public view

Rex Dalton, San Diego

Long-simmering feuds among personalities involved in volcano research are set to burst into public view next month with the publication of two books detailing a 1993 volcanic eruption in Colombia that killed six researchers.

The books give starkly different accounts of events leading up to the eruption on 14 January 1993. In *No Apparent Danger*, Victoria Bruce, a geologist and writer based in Annapolis, Maryland, criticizes the actions of volcanologist Stanley Williams of Arizona State University, who led the fateful expedition into the crater of the Galeras volcano. But in his own book, *Surviving Galeras*, written

with journalist Fen Montaigne, Williams strongly defends his decision to proceed with the expedition.

Bruce alleges that Williams ignored important seismology data that suggested Galeras may have been about to erupt. She claims that he was cavalier about safety, thereby contributing to injuries, and misrepresented his role in the incident for years afterwards. She further charges that Williams later misappropriated scientific ideas from another scientist, using them to secure a grant and publish an article with his graduate students.

Williams, who received near-fatal head and leg injuries in the Galeras explosion, denies Bruce's charges, saying he has no regrets about his actions. "The idea of me stealing science is completely untrue," he adds. But writing his book has forced Williams to acknowledge that he was not the sole survivor of the explosion, as he has repeatedly claimed in the past. "I made a mistake; I am sorry," Williams says, adding that he has apologized to the other survivors.

At the time of the explosion, Williams was conducting a volcanology workshop. This included a trek into the volcano crater while members of the Colombian media and others observed from the rim. While tests were under way, a volcanic explosion blasted out rocks, killing scientists from Colombia, Russia and Britain. Williams' skull was fractured, his leg was splintered and he nearly died before he was evacuated. He incurred a moderate brain injury.

Killer 'hiccup'

In her book, Bruce — who received a masters degree in geology from the University of California, Riverside, for studies on the Mount Rainier volcano in Washington state — contends that long-period seismology data, which reflect pressure fluctuations in a volcano, suggested that Galeras was potentially volatile at the time of the research trip. Williams ignored these seismology data, she contends.

But Williams says that appropriate significance was attached to the seismology data, and that they "were discussed in great detail" by the volcano explorers. In his book, Williams describes lying injured in the volcano, reviewing his preparations for the trip, and concluding that he couldn't have predicted the explosion. In comparison with other eruptions, the rock burst was "a hiccup", he writes. "We studied the best available data ... but Galeras behaved capriciously. I was fooled, and for that I will take responsibility. But I do not feel guilty about the deaths of my colleagues."

Bruce further charges that Williams misappropriated for publication (*Nature* 368, 135–137; 1994) research concepts on long-period seismology devised by Bernard Chouet, a seismologist with the US Geological Survey (USGS). From his office in Menlo Park, California, Chouet backed Bruce's account. "He ripped me off," says Chouet, who says he presented his idea at a November 1991 volcano workshop that was attended by Williams. "I am sorry he [Chouet] feels that way," says Williams. "He is not correct."

In dispute

The incidents described by Bruce are not the only ones in which Williams' style has brought him into conflict with other volcanologists.

Norm Banks, a retired USGS geologist, says Williams engaged in an "unprofessional assumption of authorship" of Banks' ideas and data on volcano pyroclastic lava flows in Colombia (*Geology* 20, 539–542; 1992). Banks says he never again shared any unpublished ideas with Williams. Denying Banks' charge, Williams says: "I'm sorry to hear he looks back at that negatively."

Nearly two years after the Galeras eruption, Williams was embroiled in a dispute over access to an active South Pacific volcano and his subsequent use of data from Rabaul in Papua New Guinea.

After the USGS declined Williams' request to join its relief mission, Williams obtained emergency funding from the National Science Foundation (NSF) and went to Rabaul anyway.

Christopher McKee, an Australian seismologist, was at the time head of the Rabaul Volcanological Observatory. Now at the Geophysical Observatory in Port Moresby, Papua New Guinea, he says Williams did a "snatch and grab raid" for samples for his own publication (*Science* 273, 490–493; 1996), leaving the local team feeling as if some of their "work had been stolen from them".

Wally Johnson, a geologist who was then secretary-general of the International Association of Volcanology and Chemistry of the Earth's Interior, says he wrote to the NSF complaining about Williams' trip. Johnson, who was at Rabaul, says he found Williams' presence there disruptive. But he adds that Williams "is a good scientist who does good science".

Williams says he acted appropriately in Rabaul. "I am an aggressive person," he says. "I am not eager to have US government officials control my thinking. But over the long run, I have remained active on the scene with good results."



Deadly data: were seismological readings suggesting volatility ignored at Galeras?

Trials offer way forward for Parkinson's

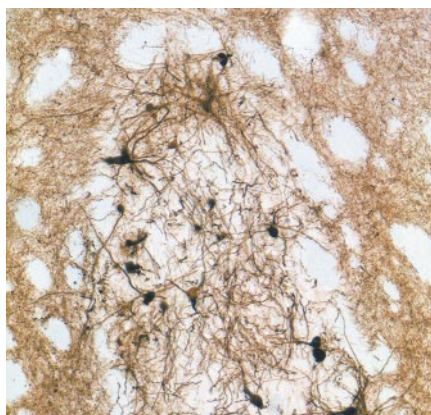
Alison Abbott

Despite widely reported negative side effects, the results of the first controlled clinical trial using transplants of fetal cells to treat Parkinson's disease are positive for the therapy's long-term prospects, researchers say.

But scientists caution that the trial's design makes it difficult to interpret. Published this month in the *New England Journal of Medicine* (344, 710–719), the trial was led by Curt Freed of the University of Colorado.

Freed transplanted precursors of dopamine-producing cells from human fetal brains into the brains of 20 patients with severe Parkinson's disease. Twenty other patients received 'sham surgery' in which no cells were transplanted. The transplants survived and produced dopamine, the lack of which causes the disease. The results prove the principle of such transplants for the first time in a controlled study, researchers say. But the reported clinical benefits were modest, and five patients developed severe side effects.

Parkinson's disease is characterized by movement disorders ranging from tremors to full rigidity. It can be treated with L-DOPA, which is converted into dopamine in the brain. But after long-term treatment, most patients start to respond unpredictably. Some become



Natural source: fetal tissue transplants can produce dopamine in patients with Parkinson's.

oversensitive to the drug and experience uncontrolled limb movements, or dyskinesias.

The idea of transplanting fetal cells is to boost dopamine levels using a more natural source — human cells. Two US teams have been funded by the National Institutes of Health to conduct controlled trials to test the technique: Freed's is the first, and the second, led by Thomas Freeman of the University of South Florida, will finish next year.

In Freed's trial, five patients developed

severe dyskinesias more than a year after the operation, leading to press reports that the trial had gone badly wrong (see *The New York Times*, 8 March 2001). But Olle Lindvall, a neurologist who pioneered the transplant technique at the University of Lund in Sweden, disputes this interpretation. "Freed's study is important because it has shown for the first time a specific graft-induced response," he says.

Freed says that the dyskinesias might have been caused by overproduction of dopamine. But Lindvall says that there is absolutely no evidence of dopamine overproduction, and that the mechanism remains totally unclear.

Freeman says that Freed's study was also hampered by its subjective primary endpoint, in which patients' progress was assessed by interview one year after surgery. "This was a mistake," concedes Freed, who believes that it underestimated the patients' true clinical improvement.

Researchers in the field agree that any future cell therapy for Parkinson's will have to use cultured cell lines to provide a better source of transplant material. Researchers are also trying to genetically engineer cell lines that will self-destruct after transplantation if severe side effects develop. ■

OLLE LINDVALL

Bush U-turns on pledge for carbon dioxide emissions

Mark Schrope

In a move that has dashed environmentalists' hopes that the new administration might accommodate their views on global warming, President George W. Bush has reversed a campaign pledge to place mandatory caps on the emission of carbon dioxide from power plants.

In a letter to four senators, Bush attributed his decision to the "incomplete state of scientific knowledge of the causes of, and solutions to, global climate change and the lack of commercially available technologies for removing and storing carbon dioxide".

Bush's action was widely criticized by environmentalists, European government officials and moderate Republicans. Sherwood Boehlert (Republican, New York), chair of the Science Committee in the House of Representatives, called it "a misguided and unjustified reversal of position".

On 14 March, the day after Bush's announcement, Boehlert's committee held a meeting on climate-change research at which all witnesses testified that greenhouse-gas emissions were contributing to global climate change. Boehlert later said: "I've got some advice for the president: pay

attention to the scientific enterprise."

Legislation to amend the Clean Air Act so as to cap carbon dioxide emissions has been proposed by Boehlert and a bipartisan group of congressmen and senators. But it is expected to fail without Bush's support.

Sean McCormack, a Bush administration official, says the administration's approach to climate change is being totally reassessed. "Other parties should not make any assumptions about what US policy will be until the review is complete," he says.

International leaders took the unusual step of commenting on Bush's decision. Margo Wallström, environmental commissioner for the European Union, said in a statement that she was concerned about Bush's characterization of the science of global climate change. And Klaus Toepfer, executive director of the United Nations Environment Programme, said in a speech in Denmark that he doubted there could be an effective global response to climate change without US leadership. ■



Evaporating hopes? Bush's decision could delay action to slow the melting of glaciers.

STILL PICTURES

Universities hit as Californian energy crisis starts to bite

San Diego California's energy crisis took its toll on the state's universities last week, as funding for three new research centres came under threat. The universities may get their revenge, however — they have teamed up to sue utility company Enron for alleged breach of contract.

A state committee scrapped \$75 million worth of funding two weeks ago for new University of California (UC) research centres devoted to biotechnology, telecommunications and nanotechnology. A separate committee restored the funding last week, but it may be cut again when the budget is finalized in May. Funding for a fourth centre, planned to specialize in information technology, is still suspended.

In a separate development, fears of escalating utility bills have led UC and the California State University to sue electricity company Enron of Houston, Texas. The universities say Enron is seeking to break the contract — an action which would prompt campus electricity bills to skyrocket, at the potential expense of teaching and research.

If the contract is breached, the universities say they would face extra electricity costs of up to \$300 million over ten years. Enron denies that it is trying to break the contract and increase university utility bills. A hearing date on the lawsuit has not been scheduled.

Lab 'deceived' scientist into working on warheads

San Diego Lawrence Livermore National Laboratory (LLNL) in California has been accused of using federal funds designated for the safety and maintenance of the United States' existing nuclear weapons to develop new designs for warheads.

LLNL computer scientist Isaac Trotts resigned from his \$85,000-a-year post last week, saying that LLNL had "deceived" him into working on enhancing nuclear weapons. Trotts, employed on a Stockpile Stewardship, claims he agreed to work at LLNL on the condition that his research didn't contribute to the development of nuclear arms.

An LLNL spokesman said he couldn't discuss Trotts's case specifically because of personnel laws, but added: "Our recruiting is honest — there is no foundation for the statements he made."

Thousands flock to garden of Eden

London The Eden Project, a UK park designed to showcase plants from around the world, opened last weekend.

The £80 million (US\$113 million) project



Home sweet biome: the Eden Project will house a vast collection of plants, some endangered.

centres on two giant greenhouses, or 'biomes', built in a disused clay pit in Cornwall, southwest England. These are stocked with plants from warm and tropical regions — a 25-metre-high waterfall in the tropical biome helps to maintain humidity. A third, roofless, area features flora from temperate regions. Eden will host doctoral researchers through collaborations with the universities of Reading, Exeter and Plymouth.

Some 7,000 people visited the park on its opening day. Disinfectant pads have been put at the entrances to allay fears of spreading foot-and-mouth disease.

Court use of blood sample 'broke confidentiality'

London AIDS researchers have expressed concerns about patient privacy, after a supposedly confidential blood sample from a research project was used as evidence in a Scottish criminal trial.

Stephen Kelly of Glasgow was found guilty last month of knowingly infecting his girlfriend with HIV. Now it has emerged that blood samples taken from Kelly during a study of needle-sharing in Scottish jails were used as evidence against him, despite assurances that all samples would be kept confidential.

Andrew Leigh Brown of Edinburgh University, who led the study, has said he will not undertake similar work in Scotland until the situation is clarified. The study was paid for by the government-funded Medical Research Council, which says confidentiality can be waived in exceptional cases.

Coral campaign gets massive funding boost

Washington The International Coral Reef Action Network (ICRAN), which campaigns to protect coral reefs, received a \$10 million grant from Ted Turner's United Nations Foundation (UNF) last week.

ICRAN — a coalition of academic, private and government organizations — intends to use the money to fund sites for demonstrating coral-reef management in the Caribbean, East Africa, East Asia and the South Pacific. Educational programmes based at these sites will provide information on conservation practices in the hope of

reversing the worldwide decline of coral reefs.

The UNF was created in 1997 to distribute the \$1 billion pledged by entertainment mogul Turner for United Nations-related projects. This is its largest environmental donation to date.

Miles from metrication, but NASA inches towards it

Washington NASA is happy to go metric, the US space agency said last week — but not when it is communicating with the public.

NASA's inspector general reviewed its use of units following the loss of the Mars Climate Orbiter in 1999. The Orbiter had received incorrect instructions because one of the teams involved had been working with non-metric measurements. The US aerospace industry generally been slow to convert to metric units.

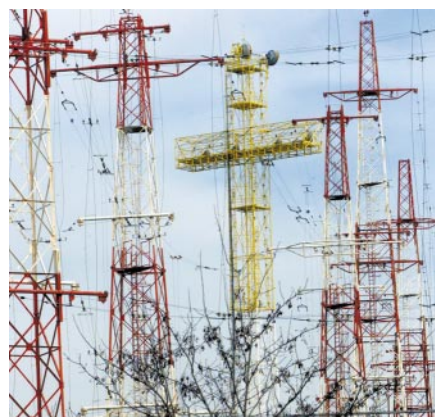
Now the inspector general has advised NASA to increase the use of metric units as far as possible. Although NASA agreed to most of the changes, it said it wouldn't tell its public affairs office to go metric, citing low recognition of metric units among the public.

Transmitter concerns could silence Vatican

Munich Radio Vatican's radio transmitters may be forced to stop spreading the word by the end of this month.

Italian environmental minister Willer Bordon claimed last week that the electrical field produced by the transmitters breaches new legal limits. The transmitters produce electric fields of 12 volts per metre, higher than the Italian legal limit of 6 volts per metre, although this is much stricter than the limit allowed by the European Union. Three Radio Vatican directors are due to face trial in October accused of causing electromagnetic pollution.

Bordon's announcement follows repeated complaints from people living near radio transmitters. Campaigners blame the transmitters for local cases of leukaemia.



Cross to bear: campaigners fear that radiation from transmitters can cause leukaemia.

'One woman is enough...'

Few women reach the uppermost rungs of Japan's scientific hierarchy. But some are now starting to challenge the system and attitudes that frustrate their career progress, as David Cyranoski discovered.

Mitiko Go, a full professor of biophysics at Nagoya University, is one of the few women who have seen first-hand the way in which senior university appointments are handed out in Japan. For Go, it was not a pretty sight. A few years ago, when the members of her Division of Biological Science discussed a female applicant for a vacant chair, one professor, now retired, said: "I'm sorry to have to say this in front of Dr Go, but one woman is enough." At the time, Go felt unable to reply. "No one would have stood by me," she says.

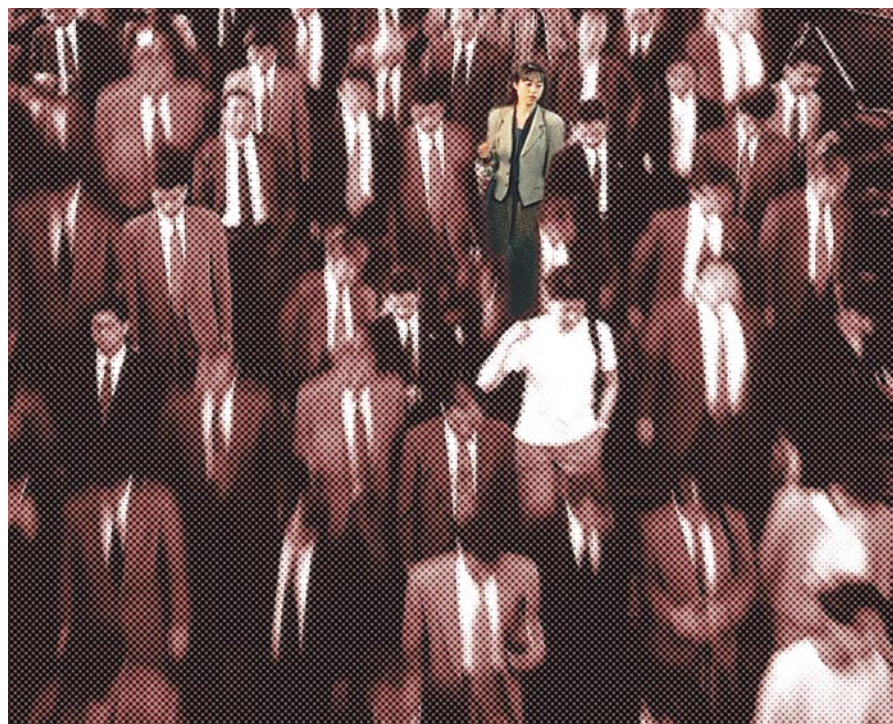
This incident is symptomatic of the discrimination — both overt and covert — that many Japanese women scientists say they experience. According to a poll conducted in November 1999 by Mariko



Mariko Kato's survey suggests Japanese women astronomers are often harassed.

Kato, an associate professor of astronomy at Keio University, 42% of women astronomers in Japan have experienced sexual harassment and about half of those claim to have "suffered emotionally or physically". But these cases rarely make it into the newspapers or the courtroom, because Japanese women fear adverse effects on their career, embarrassment and a general lack of support.

Given this background, it is no surprise that few women make it to the most senior positions in Japanese science. Women make up more than 40% of the country's overall workforce, but only 10% of the research community. Universities, with 18.4%



FRANK SPOONER

women among their research staff, have a better record than government labs (8.4%) and industry (5.1%). But even here they are overwhelmingly concentrated in the most junior positions. They may account for up to half of graduate students, postdocs and technicians, but a woman full professor is a rare

Many Japanese men feel more comfortable without women as colleagues.



Mitiko Go has seen sex discrimination in action.

thing indeed (see charts, overleaf).

Although the outlook remains bleak, some women scientists such as Go and Kato are now speaking out against these iniquities. And others are battling the odds and taking their cases to court. Last year, in a landmark decision, toxicologist Kumiko Ogoshi won a lawsuit against Nara Medical University for the harassment she faced after she set up an association of research associates — the lowest rung on the ladder of permanent academic positions — to examine the academic research system (see *Nature* 407, 935; 2000). Although Ogoshi was awarded only US\$5,000, the recognition in a court of law that harassment occurs in academia set an important precedent.

Another pending case is highlighting the plight of women scientists in temporary research posts, who often lack the security of a written contract. In 1997, a woman researcher on a three-year appointment at the National Institute for Longevity Sciences in Aichi, failed to receive a lucrative research grant that she believed she had been promised on taking the job. The grant was eventually distributed to another — male — researcher at the same institute. The plaintiff is now suing the government and her former boss for breach of promise and for harassment.

At the heart of the culture of discrimination, according to many women scientists, lies the 'koza' system, which places huge power in the hands of full professors. At Japanese universities, professors have almost complete control over funding and hiring decisions under their chair. Many of them are men with traditionalist attitudes — which means that women face entrenched, but largely hidden barriers to advancement.

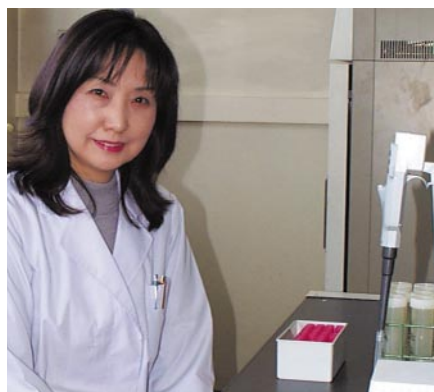
"The restraints can't be seen, but when women go to the United States, they go away, and so the women stay there," says one researcher at RIKEN, the Institute of Physical and Chemical Research in Wako, who has seen many of her friends move abroad.

Double standards

In Japan, career advancement is usually the result of being called from above, rather than exerting pressure from below. *Koza* professors are supposed to look after the future of their students, but there are widespread complaints that the level of assistance and patronage provided is radically different for men and women. One indignant woman neuroscientist relates the story of a male colleague whose research suffered during a bout of depression. His mentor rescued his career by arranging a permanent position in the lab of a personal friend. "This just would not happen with a woman," says the researcher.

Professors like hiring women postdocs, says Go, who has heard colleagues express such opinions as: "They are good hard workers, and I don't have to worry about their future." And given a recent expansion in the number of postdoctoral posts, many more women are finding themselves stuck at the bottom of the career ladder, unable to get permanent positions. Indeed, while men advance up the academic ladder, women can be told that "your husband has a good position anyway", or "you will probably be getting married soon".

If jobs were advertised openly, the situation might not be so bad. But often they are not; and even if jobs are advertised, says Sonoko Habu, a professor of immunology at Tokai University School of Medicine, "they are often decided before that". Habu recalls being actively discouraged from applying for research associate positions by colleagues of her *koza* professor — they argued that men needed the jobs to support their families. Successful candidates, says Habu, usually owe their appointments to "personal relations — who your university classmates were or who your drinking buddies are". Japan's



Beating the system: toxicologist Kumiko Ogoshi successfully sued her university for harassment.



It's not what you know: Sonoko Habu (standing centre) feels personal relations influence promotions.

culture of after-work social drinking is almost exclusively male, so women are left out in the cold.

There is also the problem of credit on scientific papers. Women are less likely to be named as first author on a paper for which they contributed the bulk of the work, says Go. If they are, those evaluating the publications during hiring or grant decisions will often assume that the woman did not do the work but was given first authorship out of kindness by the lab head. Go was blessed with an understanding mentor and a field — theoretical and biophysical analysis of protein structures — that allowed her to do most of her work alone. So she was able to publish papers under her sole authorship. "Then there is no question," says Go.

The scientific career structure in Japan also makes few concessions to women who need to juggle the demands of career and family. "Anything that goes wrong at work is seen as an indication of too much commitment to one's family," says Noriko Mochizuki-Oda, a neuroscientist currently conducting research with short-term funding from industry at the Osaka Bioscience Institute. "You lose confidence."

Women who take a few years out to start a family run up against a system in which careers are expected to be established by one's early thirties. "But that is the most difficult age for women to focus on their research," says Mochizuki-Oda. What is more, many grants are only awarded to researchers in their mid-thirties or younger. "In Japan there is still the idea of lifetime employment and seniority-based promotion all starting when you are 30," says Junko Nakanishi, a full professor in environmental risk management at Yokohama National University.

Mothers also lose out from Japan's prevailing attitudes which judge hours worked as a measure of one's worth. "Assuming the results are about the same, research by a man who stays until midnight will be more highly praised than that by a woman who

returns home in the early evening," says Yoichi Oda, Mochizuki-Oda's husband, himself a neuroscientist at Osaka University.

What's in a name?

Women fear that even getting married — with its attendant stereotype of 'housewife' — can place them at a disadvantage. And regulations from the Ministry of Education, Culture, Sports, Science and Technology exacerbate the situation. Grants-in-aid for scientific research, the greatest source of funding for academic research in Japan, will not be given to married women who try to use their maiden name. As careers often depend on name recognition, many women resent what they see as coercion to abandon their 'professional' names. Go, Kato and others have repeatedly demanded that the ministry change its policy — and according to a ministry official, the restriction will be lifted before the next round of applications this autumn.

Officialdom is slowly realizing that reforms are needed. The Science Council of Japan, which advises the Cabinet on science and technology, last July announced that it would raise its female membership to 10% by 2010. Soon afterwards, five more women



Junko Nakanishi is one of the few women to head a government-funded research centre.

members were appointed, bringing the total to seven out of 210. Government committees have been given a target to raise the proportion of women among their members to 30% by 2005. And the Japan Association of National Universities has set itself a 20% target for permanent research positions by 2010. Currently, the figure stands at 7%.

But these targets are not binding, and many women scientists doubt that they will be met. When the reorganized National Institute of Advanced Industrial Science and Technology (AIST) sought directors for its 45 new centres and institutes earlier this year, officials had in mind a 30% target. But only one female director — Nakanishi, who will head the Research Center for Chemical Risk Management in Tsukuba — was appointed. "We just could not find the qualified people," says a female AIST official. RIKEN similarly includes just one woman among its 56 chief scientists.

Pressure point

Given the slow progress, women scientists say that the pressure being exerted by prominent scientists such as Go and Kato, and by those women who are prepared to take their grievances to court, is extremely important.

The next focus of attention is likely to be on the pending action against the Aichi longevity institute. In this case, the plaintiff will argue, using evidence from e-mail exchanges, that she had been promised a lucrative Japan Science and Technology Corporation special research grant that would more than double her pay. She also claims that her division head had told her it would be possible to extend her three-year position for "a long time", and that he promised her a subsidy from his research fees.

As is common in Japan, none of this was written into an official contract. "In Japan,

Favoured in a foreign land

One small group of women seems to escape the sexual discrimination that blights Japan's research system — foreigners. "Japan can be a lot more gender-neutral than people would think," says Kathleen Rockland, an American neuroscientist who one year ago was appointed as a team leader at the RIKEN Brain Science Institute in Wako.

Foreigners in general are given a certain amount of leeway

in Japan, and it seems that foreign women are put in a special class. They are anticipated to be strong characters, and as such are treated with a degree of respect rarely accorded to Japanese women.

Suzi Jarvis is one of the few foreign research group leaders in Japan and earned her position after working for six years at the Joint Research Center for Atom Technology in Tsukuba. She believes that foreign women, as a

result of their inherent novelty, are more able to assert themselves. "I never had a problem specifically because I am a woman," she says, adding that the threat of physical harassment — at least for foreigners — is far less in Japan than, for example, in her native England. "Not everything has been easy, but wherever you go you are going to hit obstacles, and you just have to go around them, or over them, or through them."

the employee has no choice but to believe what such a high-ranking official claims to be able to offer," says Minoru Wakabayashi, a lawyer involved in the suit, which has been filed against both the division head and the Japanese government. The division head argues that he only said the grant would go to the plaintiff because he assumed no one else at the institute would apply.

Following her legal victory last year, Ogoshi continues to fight the cause of junior researchers — especially women — who are foiled in their attempts to reach the upper echelons. Last month, her organization of research associates petitioned the government to establish a non-profit group to collect data on the status of women scientists. "The fact that there are so little data collected on this issue shows that it is not considered socially significant," says Ogoshi. But there are some signs of change. The science ministry, for example, has this year started to break down data on its grant-in-aid recipients by sex.

Go, as the senior member of the Division

of Biological Science at Nagoya University, has reached the top of her profession, and was last month elected onto the university's governing senate. Although conditions at Nagoya are better than at other universities, says Go, there is still room for improvement. Go will retire in two years, and she intends to devote her remaining time in employment to increasing opportunities for women by encouraging them to apply and ensuring that they are given fair hearings for promotions and new positions.

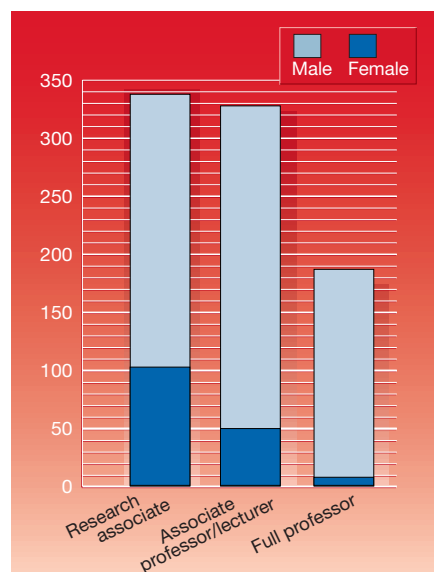
Kato, meanwhile, is turning her attention to another iniquity of the science ministry's grants-in-aid system. At the moment, the ministry only accepts applications from those holding a permanent position, which Kato says discriminates against the many women who are left stranded in temporary research posts. "It exacerbates the problem of the predominantly female researchers who are unable to get a permanent position," she says.

Go and Kato are also trying to create a more favourable environment for women through their scientific societies. Encouraged largely by Kato, the Astronomical Society of Japan established nurseries at its conferences so that mothers with young children can attend. The Biophysical Society of Japan, whose president is Go, and several other societies have now followed this lead.

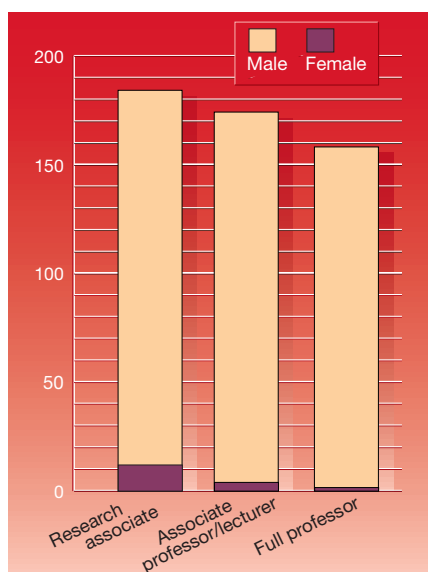
But such moves, although welcome, leave the system that frustrates the ambitions of women scientists intact. And breaking that will require a sustained and aggressive assault — something that will not be popular in Japan. During interviews for this article, several women who have serious gripes about their situation nevertheless expressed reservations about moving to a system in which, as one put it, "people are always complaining in order to get what they want".

Researchers such as Go have learnt to accept that victory will not come easily. "There is a deep psychological bent in many men in Japan," says Go. "They just feel more comfortable without women as colleagues."

David Cyranoski is *Nature's* Asian-Pacific correspondent.



Glass ceilings: the discrepancy between the sexes in Japan is clearly visible in surveys of the number of researchers in university public health/medical schools in 2000 (left) and in astronomy in 1999.



Sleepless in Seattle

A new superconductor had physicists burning the midnight oil at their get together last week. Sarah Tomlin joined in the fun.

There were some tired eyes at the meeting of the American Physical Society (APS), held last week in Seattle. A new superconductor had physicists acting like teenagers, cramming lecture theatres to see the stars of the show and staying up until the early hours to discuss the latest gossip. The session was immediately christened 'Woodstock West' — a nod to the original 'Woodstock of Physics', the 1987 APS session on the first high-temperature superconductors, which are based on copper oxide.

Centre stage this time around was magnesium diboride (MgB_2), a new superconductor that burst onto the scene this January. It was unveiled at a conference in Japan, when Jun Akimitsu of Aoyama Gakuin University in Tokyo revealed that MgB_2 loses all its electrical resistance and superconducts at temperatures of up to 39 K (ref. 1), twice the previous record for a stable metallic superconductor. The compound is also cheap and easy to make. "It's your fantasy material," says Paul Canfield, leader of an Iowa State University group, who has already made superconducting wires from the compound².

The earliest superconductors were all metallic, but needed cumbersome and expensive equipment to cool them to a few degrees above absolute zero before they lost their electrical resistance. That all changed in 1986 after the discovery of high-temperature

superconductivity in the copper oxides. By the time of the original physics Woodstock, materials such as yttrium barium copper oxide had been found to superconduct at up to 90 K. Simple metallic superconductors were almost forgotten by those searching for superconductivity at higher temperatures.

Now they are definitely back in fashion. Once news of MgB_2 's superconducting properties was out, a frenzy of e-mails and phone conversations sent the story around the physics world, culminating in last week's hastily arranged meeting.

Instant celebrity

As the word spread, physicists switched their research to concentrate on MgB_2 . "We've been working 16 hours a day, seven days a week on this," says Canfield. "It's been damn exciting fun." The latest findings were shared over the Internet via the Los Alamos archive (<http://www.arxiv.org>), a repository of preprints. Back in 1987 it was the fax machine that powered the spread of results. This time, said Princeton University's Bob Cava at the meeting, "the excitement was fanned by instantaneous communication".

If the past two months were a sprint, the 12 March evening session was a marathon. Midnight came and went before the 80 speakers had finished. More than 1,000 physicists crammed into the room, with yet more outside, to hear Akimitsu start the talks. Subsequent speakers were allowed only two minutes, strictly enforced, with one minute for questions.

Several groups described how MgB_2 could be made into wires and thin films — the first steps towards practical devices. The Iowa State researchers have found a simple way to make superconducting wires: they expose commercial boron fibres to magnesium vapour at 950 K. Canfield believes the same technique could produce superconducting magnets or other devices.

From the evidence presented at Seattle, it seems that MgB_2 behaves like a conventional low-temperature metallic superconductor — except, of course, it works at a much higher temperature. This came as a surprise to most physicists, who have struggled to work out what is going on in the copper oxides, which appear to follow a different, and largely mysterious, mechanism.

But if MgB_2 is a conventional supercon-



Hot stuff: Jun Akimitsu (left) stunned physicists when he revealed that magnesium diboride superconducts at surprisingly high temperatures.

ductor, what is it that makes MgB_2 conduct at higher temperatures? Cracking that problem could lead to new superconductors. "If we can understand what makes this structure survive at these temperatures, then it might point the way to other materials," says Canfield.

So far, attempts to improve MgB_2 's performance have been unsuccessful. For example, a tried and trusted method for improving the high-temperature copper oxide superconductors is to change their composition by chemical substitution. But when Cava's group tried to replace some of the magnesium in MgB_2 with aluminium, the superconductivity was destroyed³. It is still early days, and instead of substitution, adding a third element to MgB_2 may open up more possibilities.

The next step is to use MgB_2 to make practical devices. Researchers are cautious when talking about potential applications, in part because many of the proposed uses for high-temperature superconductivity failed to materialize. Copper oxide superconductors are hard to engineer into wires, so standard metallic superconductors such as niobium-tin and niobium-titanium are still widely used for superconducting applications such as magnets for magnetic resonance imaging instruments.

Even though MgB_2 has a moderate superconducting temperature, its low cost may make it a viable alternative to the copper oxide superconductors. Whether the excitement of the past few weeks translates into real applications remains to be seen. But the tired physicists heading back to their labs after their marathon session in Seattle are on the case.

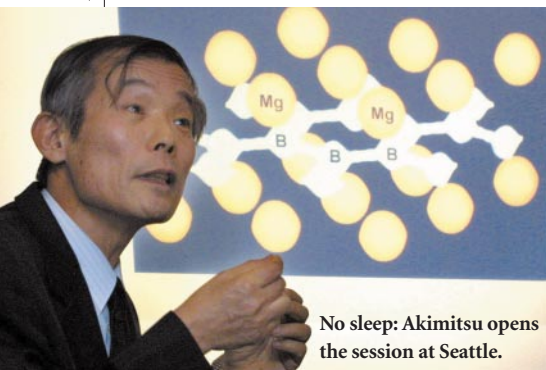
Sarah Tomlin is associate News & Views editor at Nature.

1. Nagamatsu, J. *et al.* *Nature* **410**, 63–64 (2001).

2. Canfield, P. C. *et al.* *Phys. Rev. Lett.* **86**, 2423–2426 (2001).

3. Slusky, J. S. *et al.* *Nature* **410**, 343–345 (2001).

Whether the recent excitement translates into real applications remains to be seen.



No sleep: Akimitsu opens the session at Seattle.

Action is needed now, or BSE crisis could wipe out endangered birds of prey

Sir—As pointed out in your News feature “Testing times for BSE”, cases of bovine spongiform encephalopathy (BSE) are being reported in growing numbers across continental Europe¹. Not only is the disease having an impact on human health and on countries’ economies, but it may also cause a catastrophic decline of some endangered European species of animals.

With about 30 cases of BSE detected in Spain in the past few months, the government has adopted measures aimed at arresting its spread. Since 1 January, a national law obliges farmers to incinerate all dead cows, sheep and goats, regardless of whether they are infected. From 1 March, this law has been extended to pigs, poultry and horses, despite no clear evidence for this latter group being a risk for BSE transmission. These measures are laudable from a prophylactic point of view, but they will deprive other species, some endangered, of food.

In Spain, scavenging birds have coexisted with livestock farmers for centuries². Spain now has 17,500 pairs of griffon vultures (*Gyps fulvus*), 1,200 pairs of cinereous vultures (*Aegypius monachus*), 1,300 pairs of Egyptian vultures (*Neophron percnopterus*) and 80 pairs of bearded vultures (*Gypaetus barbatus*), comprising 80–99% of the breeding pairs of vultures in the European Union. Spain also supports 80% of European red kites (*Milvus milvus*) in winter, and the entire world population — 130 pairs — of Spanish imperial eagles^{3,4} (*Aquila adalberti*).

Spanish populations have been used for captive breeding and reintroduction programmes in France, the United Kingdom, Italy, Germany and Austria, where the species are extinct or nearly so. The four species of vulture to survive 60–100% of their food from livestock carrion either abandoned or left out for them⁵. Griffon vultures alone eat about 10,000 tonnes of dead livestock per year⁴.

On 22 February, Spanish environmental groups and ecologists working on vulture conservation urged the government to consider measures to protect human health that are compatible with wildlife conservation⁴. This is not an easy task.

The first priority is that food destined for vultures must be free of BSE to avoid any risk of disease transmission to other species. But, as reported in your News feature, there is no diagnostic test yet available that reliably detects which animals are incubating the disease. Even poultry and pigs are now cautiously considered as at

risk, although transmission of BSE to chickens has not proved possible even by direct brain inoculation.

Second, research is needed to design a network of installations to provide food to scavenging birds while excluding other scavengers such as foxes. If their main food source is removed and no remedial action taken, populations will crash and decades of European efforts to conserve endangered birds will have been in vain.

Jose L. Tella

Department of Applied Biology, Doñana Biological Station, Research Centre for Environmental Science, Avda. M. Luisa, s/n, 41013 Sevilla, Spain

1. *Nature* **409**, 658–659 (2001).
2. Donazar, J. A., Naveso, M. A., Tella, J. L. & Campión, D. in *Farming and Birds in Europe* (eds Pain, D. J. & Pienkowski, M. W.) 117–149 (Academic Press, London, 1997).
3. Tucker, G. H. & Heath, M. F. *Birds in Europe: Their Conservation Status* (BirdLife Conservation Series No. 3, Cambridge, 1994).
4. Group for the Conservation of Vultures. Unpublished report (22 February 2001); <http://www.seo.org>.
5. Donazar, J. A. *The Iberian Vultures* (Reyero Editions, Madrid, 1993).

We must not be bound by anti-GM extremists

Sir—In his review (*Nature* **409**, 559–560; 2001) of Alan McHughen’s book *Pandora’s Picnic Basket*, Dick Taverne expresses surprise that the volume contains only “a cursory mention” of “the potential of transgenic crops to fight hunger and disease in the developing world”. There is an important and subtle issue therein.

McHughen is right to de-emphasize the potential advantages of gene-splicing, or genetic modification (GM), to agriculture. Whatever the benefits of GM, the likelihood of risk in the vast majority of experiments or commercial uses is so minimal that the issue of safety stands on its own. The temptation for proponents of biotechnology to emphasize benefits not only obscures the theoretical and empirical evidence of the extraordinary precision and predictability of GM and the safety of its products, but it creates a kind of logical trap. It enables opponents of GM to argue that if the ultimate benefits will be small — such as the advantages of a tomato with a longer shelf-life or a sweeter melon — we should not tolerate any risk at all of creating an invasive, weedy or toxic plant. Hence we should not permit field experiments, or should institute draconian case-by-case review of all proposals.

Although they have seldom been presented as such, the current controversies over the testing and use of GM organisms are really about academic and individual freedom — which is being systematically undermined by discriminatory and onerous regulations. If, for example, a high-

school student doing a biology project takes a packet of ‘conventional’ (but genetically improved via plant breeding) tomato or pea seeds to be irradiated at the local hospital, and plants them to investigate interesting mutants, he or she need not seek any approval from any regulatory authority. However, if the seeds have been modified by the addition of one or a few genes via gene-splicing techniques that are more precise and predictable than conventional plant breeding, the student researcher could face a mountain of paperwork and expense (to say nothing of the possibility of vandalism).

In the United States, bureaucratic requirements by the Department of Agriculture make field trials with GM organisms 10–20 times more expensive than the same experiments with virtually identical organisms modified with conventional genetic techniques.

It is irrelevant whether the purpose of crafting a new plant variety or micro-organism is to test a scientific hypothesis or a marker gene, to offer marginal improvements or “to fight hunger and disease”. Western democratic societies have long traditions of relatively unfettered agriculture research, except when *bona fide* safety issues are raised.

Traditionally, we shrink from letting authoritarian minorities dictate our social agenda. Extremists should not be allowed to dictate the terms of the GM debate.

Henry I. Miller

*Hoover Institution, Stanford University
Stanford, California 94305-6010, USA*

Jewish emigrants and German science

Sir—The News report (*Nature* **409**, 443; 2001) on my research concerning emigrants from the Kaiser Wilhelm Society stated that the Max Planck Society has “reneged on a 1948 promise to contact Jewish scientists expelled from its laboratories during the Third Reich and offer them their jobs back”.

To the best of my knowledge no such promise was ever made and there was no general policy for the reintegration of emigrants. The activities of the Max Planck Society in 1948 were directed more towards the recruitment of foreign scientific members from a small number of renowned emigrants, a symbolic gesture intended merely to restore the organization’s international contacts and reputation. To what extent personal and unofficial attempts were made to offer the jobs back is currently under investigation.

Michael Schüring

Max Planck Society, Wilhelmstrasse 44, D-10117 Berlin, Germany

Urban myths of organic farming

Organic agriculture began as an ideology, but can it meet today's needs?

Anthony Trewavas

There is a widespread belief that low-yielding organic agricultural systems are more friendly to the environment and more sustainable than high-yielding farming systems. The current aims of organic systems — maintenance of soil fertility, avoidance of pollution, use of crop rotation, animal-welfare concerns and wider environmental aspects — would be hard to quarrel with. But the rules and regulations that have to be followed to achieve these ends caused one leading organic researcher to admit that that in organic farming “there is very little science” and “this gives rise to a great deal of illogicality and confusion particularly in some areas of production”¹.

Only two principles really distinguish organic farming from other farming methods. Soluble mineral inputs are prohibited (Box 1) and synthetic herbicides and pesticides are rejected in favour of natural pesticides (Box 2, overleaf). But agriculture based on these principles results in a more costly product, mainly because of lower yields and inefficient use of land.

Organic agriculture developed from the philosophical views of Rudolf Steiner and later Lady Eve Balfour, who in the 1940s founded the Soil Association. In the United Kingdom this association licenses about 70% of organic production and sends inspectors to check that its regulations are being followed. Although its supporters assert that organic agriculture is superior to other farming methods, the lack of scientific studies means that this claim cannot be substantiated.

Conventional agriculture is a diverse set of technologies using the best available knowledge, whose ultimate goal is the safe, efficient provision of foods in abundance and at lowest price. As with all technologies, problems often arise in the practices of conventional agriculture — but rejection of a technology because of problems also means losing potential benefits.

There is a widely held belief that organic farming is environmentally superior. But although reduction in pesticide use (Box 2, overleaf) leads to higher reported levels of some insects and reported sightings of birds on organic farms (pages 121–138 of ref. 1), current synthetic pesticides are very unstable; only transient declines of most field insects are reported even at full pesticide dosage². Similarly, the lower levels of aphids observed on organic farms could well reflect lower nitrogen and protein content of organ-



Organic farming claims environmental superiority, but lack of scientific data makes this hard to verify.

ic crops, and lower yield. Expressed as a ratio of crop yield/aphid population, the difference is negligible². It is also often overlooked that some conventional mixed farming can maintain species diversity. For example, conventional mixed farming in smaller plots (providing more field margins) or farming based on the traditional ley system (for example undersowing wheat with legumes) maintains conventional yields and low costs¹. The benefits for wildlife equal those provided by organic farming but at far lower cost to the consumer¹.

Nor do organic farming practices necessarily conserve the environment. Competitive organic farmers keep their fields clear of weeds through frequent mechanical weeding — a method that damages nesting birds, worms and invertebrates — and high use of fossil fuels, which greatly increases pollution from nitrogen oxides³. A single treatment with innocuous herbicide, coupled with no-till conventional farming, avoids this damage and retains organic material in the soil surface. Similarly, although use of manure means higher, beneficial levels of earthworms in organic fields, there are numerous problems with the use of manure (Box 3, overleaf), including possible effects on human health⁴.

Use of soluble mineral salts prohibited by organic regulations is another contentious issue. The minerals taken out of farmland as food produce must balance those put back by some other means. Organic farmers typically rely on legume nitrogen fixation, rain water or mineral recycling

in the farm. The few detailed accountings suggest slow but accumulating mineral deficits, particularly of potassium and phosphate^{5,6}, in organic farmlands. Organic farms are required to try to balance manure

Box 1 Confusion of principles

Organic proponents assert that better plants are produced from minerals derived from manure breakdown, so that organic food is superior and improves human health. Hundreds of rigorous tests have failed to reveal better-tasting properties or improved nutritional value, but have consistently shown that organic produce has lower nitrate and protein content¹⁶. Conventionally farmed food seems to be better for children¹⁷, although rodents apparently favour organic food¹⁶. Overall cancer rates have dropped 15% during the era of synthetic pesticide use. Stomach cancer rates have dropped 50–60%, probably an effect of plentiful, cheap conventional fruit and vegetables¹⁸. But this may not be the whole story, because food mycotoxins from contaminating fungi (which can be controlled by specific fungicides) definitely contribute to European cancer rates — fumonisin and patulin are both reported to be higher in organic products^{19,20}, and failure to use effective fungicides on organic farms has led to these farms acting as repositories of disease^{21,22}. Organic farms may be protected from the full effects of disease outbreak because they are surrounded by conventional farms using proper fungicides.

CORBIS

and straw production with use on the farm itself. Excess organic manure or straw is thus usually not available to provide for inevitable deficits with year-to-year climate and agriculture variation. Ultimately, many organic farms can become dependent on products that are conventionally produced with inorganic minerals⁷.

Developments in the past 25 years have shown how conventional agriculture can be much more sustainable and environmentally friendly than organic farming⁸. A conventional farm can match organic yields using only 50–70% of the farmland. Excess food is being produced in Europe, so farmers are being encouraged by governments to set aside up to half of their land for fast-growing willow plantations which are then frequently coppiced and the wood used as fuel. With this novel conventional approach, now in commercial operation throughout Europe, total fossil-fuel use and carbon dioxide production are much lower than in organic farming⁹, and because of carbon recycling it is much more sustainable¹⁰. The plantation of willow trees, with its undercover of weeds, bird-nesting sites and mammal (including deer) and insect refuge, outperforms organic farms on any biological measure of environmental diversity⁸. But this practice crucially depends on the most efficient use of land for food production.

Application of any ecological approach to agriculture is fraught with uncertainty. Ecosystems are thought to maintain stability as a result of diverse species composition. Modern agriculture, with its single-crop monoculture system, is claimed by organic proponents to be inherently unstable and unsustainable. It is true that crops rapidly disappear from fallow fields as they cannot compete with weeds, but wild, stable monocultures of species such as phragmites, wild wheat, (genetically uniform) *spartina* and mangroves indicate that ecological stability is not understood¹¹. Furthermore, although mixed cropping (supposedly mimicking ecological diversity) can reduce disease, other crop combinations accelerate disease spread^{12,13}. Farms are land-management sys-

Box 2 Problems with pesticides and chemicals

Organic pesticides, it is asserted, work with nature and are environmentally unstable, unlike synthetic pesticides. About 60% of natural and synthetic chemicals are known rodent carcinogens²³, and around 20 different chemicals are used to maintain the safety of processed organic food¹.

Approved pesticides for organic farmers include

- copper sulphate, which has caused liver damage in vineyard workers, kills worms and is persistent in soil and produce (to be banned by the European Commission after 2002)
- rotenone, recently shown to induce Parkinson's disease²⁴
- *Bacillus thuringiensis* spores, causing fatal lung infections in mice²⁵.

Organic pesticides may be used more sparingly, yet more frequent treatments of crops with copper sulphate than good conventional practice have been reported on organic farms. Natural pyrethroids have to be used at much higher doses than some of the prohibited, equally unstable and much more effective synthetic pyrethroids, such as bioresmethrin.

tems maintained to produce food, in which farmer activity replaces normal ecosystem feedback controls.

In the search for a more environmentally sensitive way forward, integrated farm management combines the best of traditional farming with responsible use of modern technology¹⁴. This system integrates care and concern for the environment with safe, efficient methods of production. Detailed information on farm soil structure and field fertility is used to target minerals, and integrated pest management to control pesticides and avoid waste. But flexibility is emphasized, to take account of site-specific factors within a framework of conservation of wildlife habitat and landscape. Integrated farm management is a prime example of how to retain the benefits of technology while minimizing the problems. In contrast to organic farmers, who receive money for conversion, no financial supplement is given by

the UK government for good environmental behaviour and there is no government support to enable farmers to learn integrated farm management.

A common argument is that organic farming is 'holistic' and thus superior to reductionist 'chemical' agriculture on conventional farms. The dichotomy drawn between reductionist and holist views is, however, false and neither is superior to the other¹⁵. The organic system is really only an aggregate of regulations ensuring efficient use of resources, and as such no different from integrated farm management¹⁴. The organic community resists dissection of its system, claiming, for example, that direct comparisons of organic and conventional land are inappropriate and only the whole system can be compared². But resistance to comparison and examination invites suspicion. Any proper system can be subjected to a sensitivity analysis to identify constraints¹⁵. A genuine holistic approach emphasizes the importance of the context of the system. The flexible site-specific approach of integrated farm management uses a contextual attitude that is denied the organic farmer working under restrictive regulations.

Organic agriculture was originally formulated as an ideology, but today's global problems — such as climate change and population growth — need agricultural pragmatism and flexibility, not ideology. ■

Anthony Trewavas is at the Institute of Cell and Molecular Biology, University of Edinburgh, Edinburgh EH9 3JH, UK.

1. House of Lords Select Committee on European Communities. "16th Report: Organic Farming and the European Union" (HMSO, London, 1999).
2. MacKerron, D. K. I. *et al.* <http://www.scri.sari.ac.uk/document/AnnReps/AnRp989.pdf> (pp. 60–72).
3. Warne, J. *Farmers Weekly* 130, No. 12, 91–92 (1999).
4. Tschape, H. *et al. Epidemiol. Infect.* 114, 441–450 (1995).
5. Nolte, C. & Werner, W. *Biol. Agric. Horticult.* 10, 235–254 (1994).
6. Kaffka, S. & Koepf, H. H. *Biol. Agric. Horticult.* 6, 89–106 (1989).
7. Fowler, S. M. *et al. J. Agric. Sci.* 120, 353–360 (1993).
8. <http://www.iacr.bbsrc.ac.uk/lars/depts/csd/geneticgroup/willowresearch/twillowindex.html>
9. Bertilsson, G. Environmental consequences of different farming systems using good agricultural practices. *Proceedings of the Fertiliser Society* No. 332 (1992).
10. Addiscott, T. M. *Eur. J. Soil Sci.* 46, 161–168 (1995).
11. Wood, D. *Food Policy* 23, 371–381 (1998).
12. Matson, P. A. *et al. Science* 277, 504–508 (1997).
13. Zhu, Y. *et al. Nature* 406, 718–722 (2000).
14. <http://www.countrylife.org.uk/leaf/ICM.htm>
15. Kline, S. J. *Conceptual Foundations for Multi-disciplinary Thinking* (Stanford Univ. Press, Stanford, CA, 1995).
16. Woese, K. *et al. J. Sci. Food Agric.* 74, 281–293 (1997).
17. Schuphan, W. *Qual. Plant. Plant Foods Human Nutr.* 23, 333–358 (1974).
18. Coggon, D. & Inskip, H. *Br. Med. J.* 308, 705–708 (1994).
19. Kirchmann, H. & Thorvaldsson, G. *Eur. J. Agron.* 12, 145–161 (2000).
20. Lovejoy, S. B. Are organic foods safer? <http://www.inetport.com/~texasbot/lovejoy.htm> (1994).
21. Zwankhusen, M. J. *et al. Phytopath.* 88, 754–763 (1999).
22. Elton, R. *Norwegian J. Agric.* 10, 7–21 (1996).
23. Ames, B. M. *et al. Proc. Natl Acad. Sci. USA* 87, 7777–7781 (1990).
24. Betaarbet, R. *et al. www.nature.com/neuro/journal/v3/n12/pdf/nn1200_1301.pdf* (2000).
25. Mackenzie, D. *New Scientist* 162, No. 2188, 4 (1999).
26. Hovde, C. J. *et al. Appl. Environ. Microbiol.* 65, 3233–3235 (1999).

Box 3 Uses and misuses of manure

Soluble minerals are not used on organic farms. Although crude rock phosphate may be allowed, potassium chloride is banned; sylvanite, another form of potassium chloride, may be permitted. The main alternative mineral source for crop nutrients is animal or green manure. Manure treatment used on any mixed farm improves soil quality, but conventional crop rotation seems equally effective². Manure breakdown cannot be synchronized with crop canopy growth, as is desirable, but continues throughout the growing season. Ploughing in of legume crops (a necessary part of the organic method to build soil fertility) and continued manure breakdown leads to nitrate leaching into aquifers and waterways at identical rates to conventional farms¹. Degradation of organic material from manure in the soil produces significant amounts of nitrous oxide and methane, the most potent greenhouse gases. Manure is variable in composition, yielding unpredictable nutrition for crop growth: there is only a poor relationship between available nitrogen for crop growth and organic content of soil. Organic regulations recommend hay for animal feeding, but hay-fed animals infected with *Escherichia coli* O157 incubate this dangerous organism longer than 'conventional' animals fed with grain²⁶.

A wholly avoidable catastrophe

A reminder of the perils that pave the way to pharmacological utopia.

Dark Remedy: The Impact of Thalidomide and its Revival as a Vital Medicine

by Trent Stephens & Rock Brynner

Perseus: 2001. 228 pp. \$26

Thomas Dormandy

On Christmas Day 1956, in the West German town of Stolberg, a baby girl was born without ears. Eight months earlier her mother had complained of morning sickness. The baby's father, who worked for a local, family-run drug company called Chemie Grünenthal, had given his wife samples of the factory's promising new sedative. The baby was the first victim of a pharmaceutical disaster that, over the next 10 years, would kill or maim tens of thousands of yet-unborn infants. The catastrophe was wholly avoidable. To those who were young doctors during that decade it was an experience never to be forgotten.

Of course, medical mistakes were nothing new. Over the centuries excessive bleeding and purgation had probably killed millions. Injections of Koch's tuberculin were not only extremely painful but often activated rather than cured the disease, yet the treatment went on for years. Streptomycin, even in small doses, was the cause of auditory nerve damage leading to irreversible deafness in at least 5% of cases; the mechanism is still unknown. Suppression of the normal inflammatory response by corticosteroid therapy caused many catastrophic gastrointestinal perforations. But these treatments were attempts to deal with life-threatening illnesses or intolerable suffering. The risks were usually known. The motive was rarely financial gain.

None of this was true of thalidomide. It was a mild sedative, one of many. With deadly irony, its main selling point was its safety. It could be given in large doses without fear of poisoning. It was mendaciously claimed to be particularly well suited to alleviating morning sickness in early pregnancy, being sold over the counter in many countries. And these claims continued to be trumpeted long after the opposite was known, or at least suspected.

The motive was greed. Grünenthal's medical director, Heinrich Mückter, had served in Hitler's army, and during the Second World War had been involved in some dubious experiments on the civilian population of Kraków. His ethics and those of his scientific sidekicks and international counterparts were those of the concentration camp doctors. Claims of 'safety' were based on desultory experiments conducted in rats,



In denial: for years drug companies refused to compensate their victims.

in which the drug was harmless even in large doses; it was also ineffective. The least informed of scientists could have guessed that it was not absorbed from the animals' gut. None of these provisos was ever mentioned in the hundreds of thousands of 'therapeutic circulars' and 'personal letters' sent to medical practitioners, first by Grünenthal in Germany and then by companies selling thalidomide in other countries. By 1961 thalidomide was the best-selling sedative in Germany, five times more popular than its nearest competitor, Doriden (glutethimide).

Yet by 1961, doctors in Germany were increasingly seeing babies born with birth defects that were previously so rare that most had never even heard of them. One of the most common was tetra-phocomelia, literally 'four seal's limbs', where the babies' feet and hands were attached directly to their bodies. Cases of irreversible nerve damage in adults, sometimes leading to suicide, were also reported.

The evidence was deliberately suppressed and later vigorously denied by Grünenthal. Advertising was stepped up. International marketing agreements, shrouded in lavish hospitality, continued to be negotiated. A barrage of slander, innuendo and threats of legal action were mounted against those who tried to sound the alarm. The campaign against the German paediatrician Widukind Lenz, who established the critical period in pregnancy responsible for the damage, continued for years.

Lenz's revelations had added an extra dimension to the catastrophe. Even a single tablet taken between the sixteenth and twenty-sixth day after conception could cause dreadful deformities. Before home pregnancy tests appeared on the market, few women were certain that they were pregnant at that stage. Many stopped taking the tablets as soon as their pregnancy was confirmed. When they nevertheless gave birth to malformed infants, they often concluded that the abnormality was genetic and avoided further pregnancies.

Outside Germany, the Distillers Company in Liverpool, England, launched the drug under the name Distaval. 'Therapeutic guides' claimed that the tablets 'can be given to pregnant women and nursing mothers, without adverse effects on mother and child'. Soon the drug was also on sale in Canada, Australia and Scandinavia.

In the United States, the company Richardson-Merrell had 10 million tablets ready for distribution, and their factory was geared up to produce many more millions. The world's biggest drug market was saved from disaster by one middle-aged, middle-ranking woman employee of the US Food and Drug Administration, Frances O. Kelsey. Her job was to evaluate applications from pharmaceutical companies for a licence to market new drugs. Kelsey's scientific integrity — or obstinacy, pigheadedness or bloody-mindedness, all epithets widely bandied about at the time — rebelled against the

flimsiness of the evidence produced by the manufacturers about the safety of their product. Insisting on better safeguards, she faced threats of dismissal by her own chief. The delay proved critical: thalidomide was incontrovertibly exposed before the company was given clearance.

The man who later rightly received most credit for the exposure was William McBride, an Australian obstetrician who had prescribed the drug to several of his patients. He refused to believe the hearty assurances of the distributors that the resulting stillbirths and deformities were coincidence. His historic letter to *The Lancet* made him a national hero. (His subsequent fall after falsifying experimental results in an unrelated investigation was a classical case of hubris.)

At least 12,000 babies had by then been born deformed — miscarriages and stillbirths were never counted — and at least 5,000 grew up with terrible deformities. (The true number was certainly higher, because many minor defects, such as missing digits, several of which I know about, were never included in the count.) The damage, not only to the victims but to the guilt-ridden families, was devastating. It horrified many onlookers.

But it did not seem to trouble the makers. For years, Distillers indignantly refused to acknowledge responsibility, let alone try to compensate their victims. It was only mounting public outrage, shared by a number of large corporate shareholders (but *not* by successive governments) and fuelled by the campaign of Britain's *Sunday Times* newspaper Insight team, that eventually forced them to part with a fraction of their profits. The critical step was the decision of the newspaper's editor Harold Evans to appeal to the European Court of Human Rights against the rulings of English courts which blocked the publication of the most damning evidence. No complicitous chairman, director of research or head of advertising in any country was ever held to account for the greatest organized crime in the Western world since the Second World War.

Is it surprising that this chilling tale has had to wait decades to be told in book form? Perhaps not. The web of misinformation and threats of legal action must have deterred many. In the event, this short but brilliant book, jointly written by historian–novelist Rock Brynner and scientific researcher Trent Stephens, was worth waiting for. The authors follow up their account of the disaster with an excellent summary of the partial resurrection of the drug as a potentially valuable remedy in severe leprosy, in a number of autoimmune diseases and perhaps in malignancies such as multiple myeloma. The one question they do not ask — understandably, as the book is a factual history — is: could it happen again? The answer is surely yes.

Pharmaceutical conglomerates have become bigger, richer and more powerful since the 1960s. These characteristics are not noted antidotes to the single-minded pursuit of profit. Individuals such as Frances Kelsey have been replaced in many countries by large, often elephantine, bureaucracies. They are more, rather than less, likely to miss atypical threats. Most importantly, public demand for chemical cures for all ills of the mind and body is even more clamorous today. If there is hope, it lies in meticulously researched books such as *Dark Remedy* to remind us of the perils that pave the way to pharmacological utopia. ■

Thomas Dormandy is at 16 St Albans Road, London NW5 1RD, UK.

..... Viral activities from a quantitative slant

Virus Dynamics: Mathematical Principles of Immunology and Virology

by Martin A. Nowak & Robert M. May
Oxford University Press: 2000. 237 pp.
£40, \$70 (hbk), £19.95, \$34.95 (pbk)

Allan Randrup Thomsen

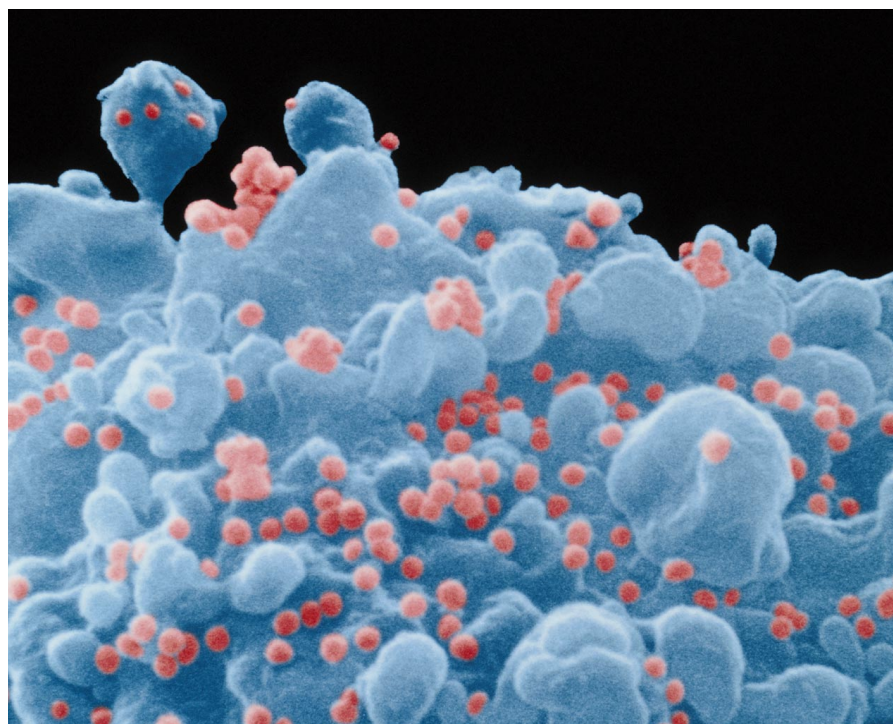
The final decades of the last century produced considerable information on the molecular biology of the immune system and its role in antiviral defences. Compared to this detailed knowledge of qualitative phenomena, our understanding of the quantitative aspects of the normal immune

response is much more incomplete. Furthermore, our insight into the dynamic interactions within the immune system following a viral challenge as well as that between populations of viruses and the immune system's lymphocytes is still very limited.

Virus Dynamics aims to close this gap by taking a mathematical approach to these issues. It views the interaction of lymphocytes and viruses in the light of population biology, somewhat analogous to a predator–prey situation. From this starting point, the authors examine important aspects of viral behaviour and the opposing immune response. They establish mathematical models that can be used not only to describe the basic dynamics of virus and immune response, but also as tools for analysing central aspects of viral behaviour in the infected host — the evolution of quasispecies, drug resistance and antigenic variation (including consequences for the evolution of the immune response in the infected host).

Generally, the models take their starting point from actual experimental observations and/or are applied to relevant viral infections, with much emphasis on HIV. This, and the fact that discussion of the mathematical models focuses on the biological implications, clearly increases the book's value and should help to build a bridge between the world of the theoretical biologist on the one hand and that of the experimentalist — and even the clinician — on the other.

The book is well written and, helpfully, gradually adds layers of understanding,



Invasion: scanning electron micrograph of the surface of a lymphocyte (blue) infected with HIV.

chapter by chapter. In this way, the models are made increasingly more realistic but also more complicated. Each chapter starts with a summary of its content and most important conclusions, and thus prepares the reader for what is subsequently presented in detail. This approach is a great advantage, for anyone not fresh from college or not using mathematics regularly will find that the mathematical deductions rapidly become complicated. It should be added, though, that the most elaborate and complicated mathematics is contained in two appendices that need not be read by the average reader for a meaningful understanding.

Experimentalists tend to see mathematical models as futile exercises carried out by mathematical nerds. The authors have succeeded in demonstrating that this is not the case, and that mathematical analysis constitutes an important additional tool for interpreting experimental data and designing future studies. The authors have also been largely successful in describing their subject in a manner that will be understandable to most people involved in studying virus–host interactions. I use the phrase “largely successful” because mathematics is the language employed in theoretical analysis, and by necessity the book therefore contains many mathematical deductions that may make it difficult reading for someone unfamiliar with that language. But I hope this does not scare off potential readers, since the book is an excellent introduction to a field that has the potential to advance substantially our understanding of the complex interplay between virus and host. ■

Allan Randrup Thomsen is at the Institute of Medical Microbiology & Immunology, University of Copenhagen, The Panum Institute, 3C Blegdamsvej, DK-2200 Copenhagen N, Denmark.

A honey-pot of knowledge

The Bees of the World

by Charles D. Michener

Johns Hopkins University Press: 2000. 913 pp. \$135, £87

Christopher O'Toole

Wherever two or more bee specialists are gathered together, sooner or later the conversation turns to “What is Mich up to now?”. The name of Charles D. Michener is synonymous with bee taxonomy and biology.

Published in Michener's eighty-third year, *The Bees of the World* is the product of more than half a century's research. At the University of Kansas at Lawrence, Michener has built up a school of bee taxonomy that is



Busy: as principal pollinators of natural vegetation and crops, bees occupy a key ecological position.

second to none. And Michener's taxonomy has always been informed by behavioural and biological studies: in addition to a huge body of taxonomic publications there is an equally impressive corpus devoted to the nesting biology and floral relations of bees and the evolution of social behaviour.

The book reflects this broad approach. There are seven chapters on various aspects of bee biology, including one on sociality. Three chapters deal with anatomy and five look at, among other things, the evolution of bees from sphecoid hunting wasps, and the higher classification and phylogeny of the insects. All topics are well illustrated by line drawings and both monochrome and colour photographs. There is an excellent 62-page bibliography that includes references to the nesting behaviour and floral relationships of bees as well as their taxonomy.

The bulk of the book (68 chapters!) is devoted to the classification and identification of bees down to the level of genus and subgenus, and Michener provides keys at both levels that are regionally based. Some of these are translations and/or reproductions of keys written by other workers, and it is extremely useful to have all of this gathered together in one volume. Inevitably, though, such an approach introduces some variation in quality. Thus, the translation of Warncke's key to the subgenera of Palaearctic *Andrena* includes many subgenera of rather dubious distinctness. Indeed, the subgenus is a highly subjective concept and many would prefer to use informal devices within a genus such as the ‘species group’, which has the advantage of not adding to the burden of formal nomenclature.

Phylogenetic purists might blanch at Michener's willingness to use paraphyletic



taxa. But he never does so without acknowledging the fact and makes a plausible case for accepting such situations as provisional and likely to be modified by more detailed future analyses. Thus, one quickly realizes that the spirit in which this book is written is one of debate and stimulus for future work: Michener always draws attention to problematic areas and offers alternative classifications to the one he happens to favour as a reasonable interpretation of presently available evidence.

Such an approach will doubtless stimulate further work, especially because *The Bees of the World* is the very first one-stop reference for the identification of bees at the generic level. In an ideal world, postgraduate students should be queuing up to start monographic revisions of genera which will allow, among other things, the identification of bees to the species level. But we don't live in an ideal world, at least not in the United Kingdom and much of Europe. Here, the research-funding culture is not sympathetic to alpha-taxonomy, that is, working to produce methods of identification at the

species level. To the extent that systematics is funded at all, the bulk of investment is in phylogenetic, molecular and DNA studies, which, although important in themselves, are of little practical use to the field ecologist, who wants to identify his material at the species level.

The problem is particularly acute for bees: pollination biologists and ecologists are ill-served by the current state of bee taxonomy, and we are in the bizarre situation whereby our understanding of pollination biology is growing faster than our ability to identify bee species.

This is a serious matter, because bees are not just another bullet point in the biodiversity crisis: as the principal pollinators of both natural vegetation and food crops, they occupy a key ecological position and, as such, are themselves vital natural resources. If we are to conserve and manage them sustainably and understand the network of bee-plant relationships on which life on Earth depends, we must be able to identify them.

Michener's magnificent book will do much to ameliorate this situation by raising the profile of bee taxonomy, and will, I am sure, stimulate more research. The book should appeal not only to entomologists interested in bees, but also to ecologists looking for an overview of bee biology. Moreover, it will be invaluable to pollination biologists who need some kind of fix on which bees are doing which things to their plants. And as far as bees are concerned, an identification to the level of genus unlocks much information on behaviour and nesting biology. ■

Christopher O'Toole is at the Bee Systematics and Biology Unit, Hope Entomological Collections, Oxford University Museum of Natural History, Parks Road, Oxford OX1 3PW, UK.

Science in culture

Suggestive signatures

Ken McMullen at the "Signatures of the Invisible" exhibition.

Martin Kemp

"I do not know if we are able to convey the beauty of this world and its extreme strangeness in order to start a fruitful dialogue between artists and scientists. Is it possible to bring this strange world back to earth through pictures or films or ...?"

This question was put to me by Hans Drevermann, virtuoso of visualization and representation of nuclear events at CERN, the European Laboratory for Particle Physics, in Geneva (see *Nature* 405, 886; 2000). Its immediate context is the "Signatures of the Invisible" exhibition, which has resulted from the collaboration of CERN scientists and 11 international artists under the aegis of the London Institute. Drevermann's "strange world" is populated by electrons, protons, neutrons, muons, quarks and suchlike, which are inaccessible to our normal powers of spatial and temporal modelling. How can this Schrödingerian world of paradox, indeterminacy, probabilities, logical contradictions and oscillating transformations be captured by the sensate media to which the artist is ineluctably wedded?

Perhaps we may take comfort from the paradoxical nature of art itself — how, for example, patches of pigment on a flat surface are transformed into a speaking likeness of Rembrandt or a landscape by Constable. Perhaps the artist's tools of illusion, allusion, evocation, imaginative inference and metaphor might serve to construct worlds that do not aspire to direct representation of weird happenings in particle physics, but rather stand as suggestive analogues to the mental pictures formed by physicists.

One strategy, among many in this stimulating show, is that adopted by its ringmaster, Ken McMullen, in his video-work, *Lumen de Lumine*. The idea sounds simple, even banal. In a black space, a red-dressed woman, swaying with hands high above her head, whirls a bright light on a chord in a wide arc around the fulcrum of her body. The maximum outer circumference of the light's orbit corresponds to a shallow circle of metal on the ground. The resulting video image is paired, mirror-wise. The idea begins to assume

resonance when we realize that the dark chamber belongs to the old nuclear reactor and the metallic ring is a cross-section of the abandoned accelerator. But the physical simplicity of the set-up still seems too earth-bound in the face of the awesome complexity of particle physics.

However, clustered around this apparent simplicity are a series of rich metaphorical and perceptual dimensions. The reversed images consciously evoke the notion of matter and anti-matter, and perpetually threaten collisions that only materialize in occasional close-ups. The swishing light, whirled by its human agent, pursues almost identical yet unpredictable paths, progressively decaying as fatigue sets in. 'Fatigue' and 'decay', we may note, are just two of the human and organic metaphors that permeate the language of engineering and physics.

As observer, we can discern five different traces of the rushing light in each scene. The first, of course, are the primary paths, witnessed through persistence of vision. But we can also focus on the elliptical glares from the moderately shiny floor, on the racing gleams on the metal rings, on the fragmentary flickers from some remaining equipment at the back of the rooms, and, most surprisingly, on four, paired secondary bright-spots that oscillate back and forth over short tracks on either side of the vertical divide. If we saw only one of these traces, how well would we reconstruct the physical set-up?

By analogy, the physicist detecting the behaviour of the most elusive particles is in the position of an observer who can only see the oscillating bright-spots, which do not in themselves obviously declare the orbiting circularity of the original source. At best, the physicist might have access to two of the secondary phenomena, say the oscillating tracks and the flickering fragments at the rear. Perhaps a brilliant visualizer could hypothesize the implicate order behind the observed phenomena and the nature of the physical set-up.

I am not imputing to McMullen the direct and conscious intention of spelling out these perceptual-cognitive dimensions. Rather, they are what arise when an artist creates something that induces complexity from simplicity, presenting the work as a field for imaginative observation. In a context such as CERN, the field could hardly be richer for any artist concerned with the business of seeing and knowing. ■

Martin Kemp is in the Department of the History of Art, University of Oxford, 59 George Street, Oxford OX1 2BE, UK.

"Signatures of the Invisible" is at the Atlantis Gallery, 91 Brick Lane, London E1, until 29 March. It will also be at the Centre d'Art Contemporain in Geneva (January–March 2002) and the Gulbenkian Modern Art Centre in Lisbon (October 2002–January 2003). Further showings are planned in Rome, Paris and Stockholm.



Chemical analysis

Language reform played an integral role in the development of a discipline.

Bernadette Bensaude-Vincent

Science is often viewed as an esoteric knowledge associated with powerful and obscure practices. This is especially true of chemistry. Chemical names and formulas, though of obscure meaning to most people, refer to tangible and edible products which can poison people or relieve their pains. Who dared to create an artificial vocabulary to replace the words used in everyday life?

The easy answer is: Antoine Lavoisier did, in 1787. At the birth of modern chemistry, according to this widely held view, the founder had to wash away a plethora of names derived from alchemy to create a logical system of nomenclature. Without being absolutely wrong, this cliché distorts history.

Lavoisier did not set out to reform the language by himself. This project was initiated by the lawyer and chemist Louis-Bernard Guyton de Morveau, who established a set of basic principles. Nomenclature should reveal “the nature of things”, he said. Simple substances should have simple names evoking their most characteristic property. Compound names should express the composition of chemical compounds. Greek etymologies should be used in preference to Latin.

Furthermore, the initiative to reform the language of chemistry preceded the chemical revolution. Eighteenth-century chemists, using new and improved analytical procedures that allowed them to distinguish between substances, repeatedly complained either that one name referred to various substances or that various local names were used for one single substance. Moreover, names were needed for newly identified substances such as cobalt and vanadium, named after figures from northern European mythology.

But the increasing number of substances was not the only driving force behind the construction of a new language. Chemists were also following the example of natural scientists, who had introduced a systematic nomenclature into botany. They were further inspired by the Enlightenment philosophers' quest for a rational and universal language.

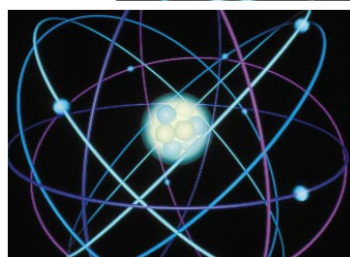
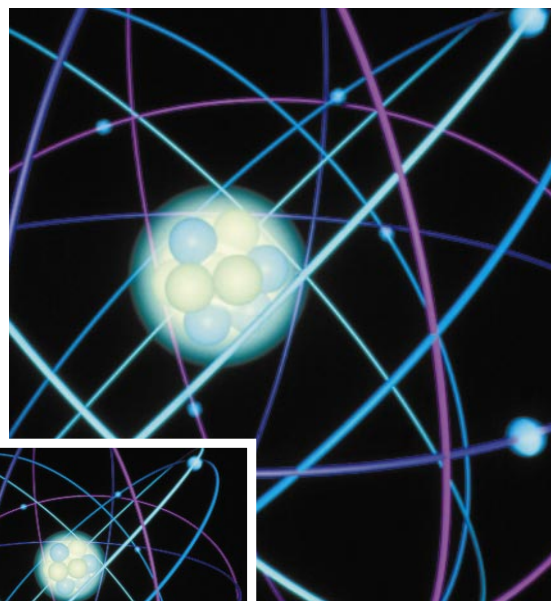
Guyton started his attempt to reform the language in 1782 and submitted his project to the Paris Academy of Sciences in January 1787. There he encountered a fierce debate over the existence of ‘phlogiston’, a principle that explained combustion and reduction in the opposite way to Lavoisier. Most chemists at the time believed in phlogiston, but Guyton was converted to Lavoisier's theories and revised his work with the help of Lavoisier,

Claude Berthollet and Antoine Fourcroy. Four months later it was published with no mention of phlogiston but with new words such as ‘oxygen’ — from Greek words meaning ‘acidifying principle’ — stemming from Lavoisier's controversial belief that all acids contained oxygen.

Lavoisier, a disciple of the philosophy of language developed by Etienne Bonnot de Condillac, argued that the new language mirrored nature and that it was a sure and definitive method because it followed “natural logic”. Surprisingly, despite numerous objections raised by contemporary chemists, the new language was adopted all over Europe during the next two decades. But adoption of the French nomenclature did not always mean conversion to Lavoisier's chemistry and philosophical commitments. The reform fulfilled a long-felt need among the chemical community and it came at the right moment as the teaching of chemistry was itself developing.

The reform of language was an integral part of the formation of the autonomous discipline of chemistry. It also contributed to the subordination of pharmacy to chemistry and, more broadly, to the redefinition of chemical arts as applied chemistry. The new language, forged by academic chemists, separated many users of chemical substances from their own traditions. Pharmacists, dyers, glass-makers and manufacturers were not really happy with the substitution of compositional terms for the terms they had used expressing colours, odours or medical properties. The new language ignored the chemists' senses, banished all reference to geographical origins or the discoverers of the substances, and imposed an analytical quantitative logic.

This logic proved to be a good method of naming over time. But the principles of the system were never strictly applied in the nineteenth century. Oxygen should have been renamed when Humphry Davy established that many acids do not contain it. But custom prevailed over the imperative of systematization. Colours and odours were restored after the discovery of chlorine and iodine, named from the Greek for ‘yellowish-green’ and ‘violet’, and bromine — from the Greek for ‘stink’. In the realm of organic chemistry, the medical properties still provided names such as morphine, named after Morpheus, god of dreams. Geographical origins resurfaced with benzene, named after *Styrax benzoin*, a



Oxygen: inspired a new nomenclature.

SPL

The new language ignored chemists' senses and imposed quantitative logic.

tree native to Sumatra and Java. Nationalism also infiltrated chemical nomenclature with scandium, germanium and polonium. In brief, it seems that the systematization imposed by four chemists, who acted as legislators in the name of rationality, remained an ideal often contradicted by practice.

This tension between the ideal of a systematic language based on general principles and the constraints of daily use remains a major characteristic of twentieth-century chemical nomenclature. Reforming the language is no longer a revolutionary enterprise conducted by four ambitious chemists in four months. It is the task of a permanent commission called the International Union of Pure and Applied Chemistry (IUPAC). New rules are formulated from time to time, as part of everyday science. But so great is the difficulty of keeping up with systematic names for complex compounds that reformers have renounced any ambition of submitting the molecular world to systematization. The ideal of a rational language, a mirror of nature, gave way to the imperative of standardization. Time will tell us whether this more modest attitude will improve the public understanding of chemistry. ■

Bernadette Bensaude-Vincent is in the Department of Philosophy, Université Paris X, 200 Avenue de la République, 92001 Nanterre Cedex, France.

Laws of form revisited

Michael Denton and Craig Marshall

Before Darwin, most biologists adhered to a platonic model of nature. This implied that the biological realm consisted of a finite set of essentially immutable natural forms that, like inorganic forms such as atoms or crystals, are an intrinsic part of the eternal order of the world. Just as, today, we account for the form of atoms and crystals by a set of physical laws or 'constructional rules', so pre-darwinian biologists sought to account for the origin of biological forms in terms of a set of generative physical laws often referred to as the 'laws of form'.

For many biologists today, platonic biology is an anachronism irretrievably laid to rest, and the idea that biological forms might be intrinsic features of nature generated by physical laws is treated with incredulity. However, recent advances in protein chemistry suggest that at least one set of biological forms — the basic protein folds —

is determined by physical laws similar to those giving rise to crystals and atoms. They give every appearance of being invariant platonic forms of precisely the type that the pre-darwinian biologists were seeking.

Protein folds, the basic constructional units of proteins, each consist of a folded chain of between 80 and 200 amino acids. Some proteins consist of a single fold, but most are a combination of two or more. During the 1970s, as the three-dimensional structure of an increasing number of folds was determined, it became apparent that the folds could be classified into a finite number of distinct structural families containing a number of closely related forms. The fact that protein folds could be classified in this manner provided the first line of evidence that the folds might be natural forms.

Further evidence that the folds do indeed represent a finite set of natural forms is provided by detailed structural studies carried out over the past two decades which have revealed that the structure of the folds can be accounted for by what amounts to a set of 'constructional rules' governing the way that the various secondary structural motifs, such as α -helices and β -sheets, can be combined and packed into compact three-dimensional structures. One is inevitably reminded of the atom-building rules governing the assembly of subatomic particles into the 92 atoms of the periodic table.

Consideration of these 'constructional laws' suggests that the total number of permissible folds is bound to be restricted to a very small number — about 4,000, according to one estimate. Confirmation that this is probably so is provided by a different type of estimate, based on the discovery rate of new folds. Using this method, Cyrus Chothia of Britain's Medical Research Council estimated that the total number of folds utilized by living organisms may not be more than 1,000. Subsequent estimates have given figures of between 500 and 1,000. Whatever the final figure, the fact that the total number of folds represents a tiny stable fraction of all possible polypeptide conformations, determined by the laws of physics, reinforces the notion that the folds, like atoms, represent a finite set of built-in natural forms.

The robustness of the folds offers another clue. The fact that the folds can retain their native conformations in the face of multiple different sorts of short-term deformations caused by the molecular turbulence of the cell, and in the face of extensive, long-term evolutionary changes in their amino-

Protein folds

"Protein folds found in nature represent a finite set of built-in, platonic forms. Protein functions are secondary adaptations of this set of primary, immutable, natural forms."

acid sequences, is precisely what would be expected if they are natural forms, specified by physical law. Again, the fact that the same fold can be specified by many different, apparently unrelated amino-acid sequences, suggesting multiple separate discoveries during the course of evolution, is further evidence that the folds are intrinsic features of the order of nature. Finally, the fact that in many cases the same fold is adapted to very different biochemical functions is precisely what would be expected if protein functions are secondary adaptations of a set of primary, immutable, natural forms.

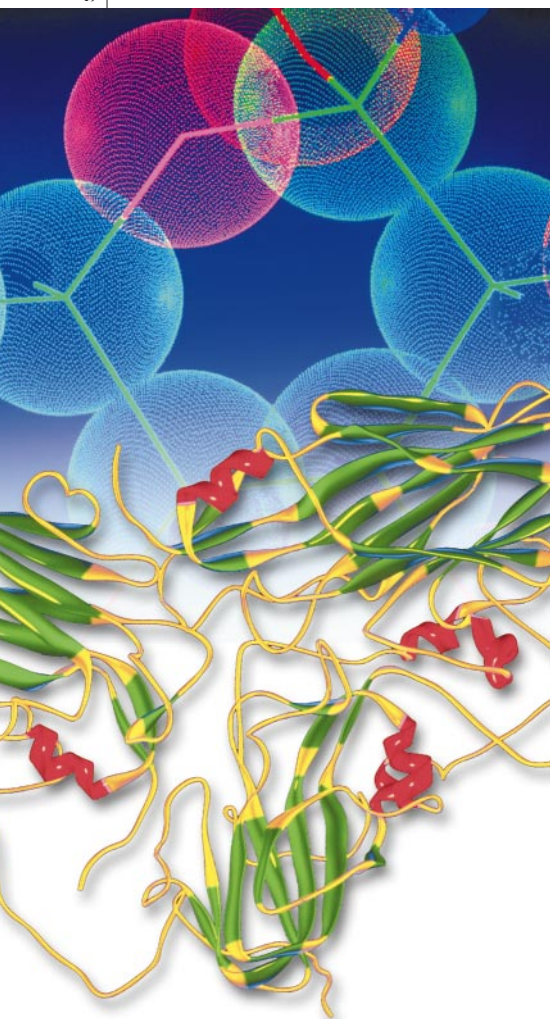
If forms as complex as the protein folds are intrinsic features of nature, might some of the higher architecture of life also be determined by physical law? The robustness of certain cytoplasmic forms, for example the spindle apparatus and the cell form of ciliate protozoans such as *Stentor*, suggests that these forms may also represent uniquely stable and energetically favoured structures specified by physical law.

If it does turn out that a substantial amount of higher biological form is natural, then the implications will be radical and far-reaching. It will mean that physical laws must have had a far greater role in the evolution of biological form than is generally assumed. And it will mean a return to the pre-darwinian conception that underlying all the diversity of the life is a finite set of natural forms that will recur over and over again anywhere in the cosmos where there is carbon-based life.

Michael Denton and Craig Marshall are in the Department of Biochemistry, University of Otago, PO Box 56, Dunedin 9001, New Zealand.

FURTHER READING

- Chothia, C. One thousand families for the molecular biologist. *Nature* **357**, 543–544 (1992).
 Chothia, C. & Finkelstein, A. V. The classification and origins of protein folding patterns. *Ann. Rev. Biochem.* **59**, 1007–1039 (1990).
 Lindgard, P. & Bohr, H. How many protein fold classes are to be found? in *Protein Folds* (eds Bohr, H. & Brunak, S.) 98–102 (CRC Press, New York, 1996).
 Kirschner, M., Gerhart, J. & Michison, T. Molecular 'Vitalism'. *Cell* **100**, 79–88 (2000).
 Orengo, C. A., Jones, D. T. & Thornton, J. Protein superfamilies and domain superfolds. *Nature* **372**, 631–634 (1994).



Another face in our family tree

Daniel E. Lieberman

The evolutionary history of humans is complex and unresolved. It now looks set to be thrown into further confusion by the discovery of another species and genus, dated to 3.5 million years ago.

Until a few years ago, the evolutionary history of our species was thought to be reasonably straightforward. Only three diverse groups of hominins — species more closely related to humans than to chimpanzees — were known, namely *Australopithecus*, *Paranthropus* and *Homo*, the genus to which humans belong. Of these, *Paranthropus* and *Homo* were presumed to have evolved between two and three million years ago^{1,2} from an early species in the genus *Australopithecus*, most likely *A. afarensis*, made famous by the fossil Lucy.

But lately, confusion has been sown in the human evolutionary tree. The discovery of three new australopithecine species — *A. anamensis*³, *A. garhi*⁴ and *A. bahrelghazali*⁵, in Kenya, Ethiopia and Chad, respectively — showed that genus to be more diverse and widespread than had been thought. Then there was the finding of another, as yet poorly understood, genus of early hominin, *Ardipithecus*, which is dated to 4.4 million years ago⁶. And earlier this year, a French team claimed to have discovered another (albeit controversial) candidate for the oldest known hominin, *Orrorin tugenensis* (ref. 7; to be discussed here next week). To those of us who are interested in reconstructing the evolutionary history of our species, these discoveries have been fun, if a little bewildering.

The confusion (and enjoyment) now looks set to increase still further. On page 433 of this issue⁸, Leakey and colleagues describe what they assert to be a new genus and species of early hominin, *Kenyanthropus platyops*. The species, whose type specimen is a spectacular partial skull, called KNM-WT 40000 (Fig. 1), was found at the site of Lomekwi on the western side of Lake Turkana in northern Kenya. The hominin bones discovered there include more than 30 skull and dental fragments, two of which have been assigned to *K. platyops*. (The other fragments have not yet been assigned to any genus or species.) These fossils were all found in deposits reliably dated to between 3.5 million and 3.2 million years ago. The other mammalian species found at Lomekwi suggest that, during this period, the site was part of a complex mixture of grassland and wooded habitats, not unlike other roughly contemporary sites such as Laetoli (Tanzania) and Hadar (Ethiopia), where remains of *A. afarensis* have been found.

Is the authors' claim⁸ — that the fossils



Figure 1 Two fossil skulls from early hominin species. Left, KNM-WT 40000. This newly discovered fossil is described by Leakey *et al.*⁸. It is judged to represent a new species, *Kenyanthropus platyops*. Right, KNM-ER 1470. This skull was formerly attributed to *Homo rudolfensis*¹, but might best be reassigned to the genus *Kenyanthropus* — the two skulls share many similarities, such as the flatness of the face and the shape of the brow. However, they are clearly different species, as *K. platyops* had a significantly smaller brain.

represent a new species and genus — likely to be true? The first part is easier to answer; KNM-WT 40000 is almost certainly a new species. The fossil has a dizzying mosaic of features. None of these characteristics is in itself new. But the combination of features is not found in any other known species, and would be hard to explain even if the species were remarkably diverse, with considerable morphological differences between sexes. The fossil resembles chimpanzees and one of the australopithecine species, *A. anamensis*, in having a small earhole. And it shares many other features of primitive hominins with *A. afarensis* and *A. anamensis*, such as cheek teeth with thick enamel, a small brain the size of that of a chimpanzee, and flat nasal margins.

But the fossil's face also has several important 'derived features' (defined as those not present in the closest known ancestor) that unequivocally distinguish it from *A. anamensis*, *A. afarensis* and *A. africanus*. These include an anterior origin for the root of the cheekbone arch on the upper jaw; the existence of a flat plane beneath the nose bone (and so the appearance of a flat face); and a tall cheek region. KNM-WT 40000 also differs from *A. garhi* in a number of ways: for example, the postcanine teeth and brow of the skull are smaller in KNM-WT

40000. The skull also lacks most of the derived features of *Paranthropus*, with a few exceptions such as the presence of three roots in the upper premolars. And, most interestingly, KNM-WT 40000 has a small cranial capacity but otherwise much in common with the famous KNM-ER 1470 fossil (Fig. 1), which is generally referred to as *Homo rudolfensis*¹. These similarities are mostly in the face, and include the flat plane beneath the nose bone, the tall, vertically oriented cheek region, and the lack of a depression behind the ridge of the brow.

A harder problem is whether KNM-WT 40000 belongs in a new genus. The difficulty is that, ideally, a genus should reflect a lineage that is unique, in terms of both its adaptations to a given environment and its relationships with other genera⁹. At present, it is hard to believe any reconstruction of hominin relationships because of the abundance of independently evolved similarities in the hominin fossil record. The complex mosaic of features seen in the new fossil will only exacerbate the problem.

Yet Leakey *et al.*'s proposal⁸ to erect a new genus, *Kenyanthropus*, for the fossil is attractive, for the simple reason that none of the other possible solutions seem feasible. First, the species does not fit comfortably in the diagnoses of any existing genus, whether

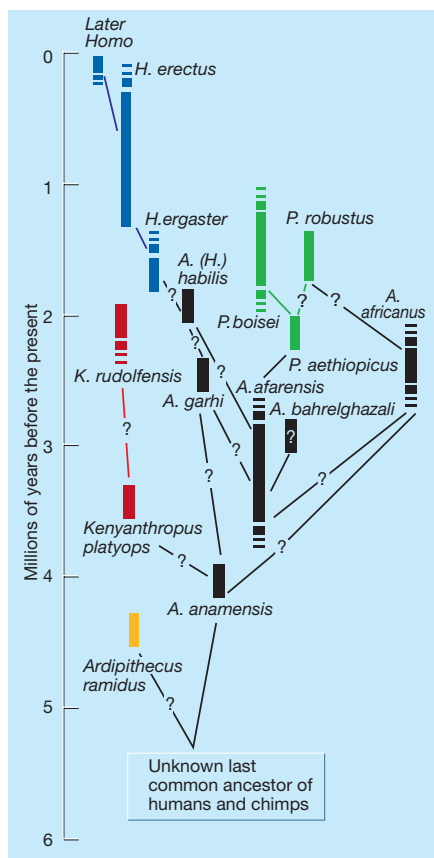


Figure 2 Possible evolutionary relationships of the hominins, indicating the five major genera, with *Kenyanthropus* in red, *Homo* in blue, *Paranthropus* in green, *Australopithecus* in black and *Ardipithecus* in yellow. Question marks indicate hypothetical or conjectural relationships; horizontal bars indicate uncertainty in the species' temporal spans.

Ardipithecus, *Australopithecus*, *Paranthropus* or *Homo*. Second, classifying the species as *Australopithecus* would also create problems because of the many derived features that KNM-WT 40000 shares with *H. rudolfensis*, which, in turn, shares other derived features with *Paranthropus*¹⁰. Species in a genus should be 'monophyletic' (have a single origin), but these similarities would probably render current definitions of *Australopithecus* species non-monophyletic. A third way round the problem might have been to dump the whole lot, including *Paranthropus* and early *Homo*, into *Australopithecus* for the time being. The advantage of this solution — my favourite of the three — is that it avoids non-monophyletic genera. But, as Leakey *et al.* point out, this scheme would render *Australopithecus* a confusing 'garbage can' genus. Hence the new genus, *Kenyanthropus*.

The nature of *Kenyanthropus platyops* raises all kinds of questions, about human evolution in general and the behaviour of this species in particular. Why, for example, does it have the unusual combination of small cheek teeth and a big flat face with an anteriorly positioned arch of the cheek bone?

All other known hominin species with big faces and similarly positioned cheek bones have big teeth. I suspect the chief role of *K. platyops* in the next few years will be to act as a sort of party spoiler, highlighting the confusion that confronts research into evolutionary relationships among hominins (Fig. 2).

The confusion is in part a testament to the intense, successful fieldwork efforts that have almost doubled the number of recognized hominin species over the past 15 years. We can now say with confidence that hominin evolution, like that of many other mammalian groups, occurred through a series of complex radiations, in which many new species evolve and diversify rapidly. It seems that between 3.5 and 2 million years ago there were several human-like species, which were well adapted to life in different environments, although in ways that we have yet to appreciate fully. But these radiations bring with them systematic headaches, because they make it hard to work out where new species fit in by using standard information

from skull and teeth fossils. A challenge for the next decade will be for skeletal biologists, palaeontologists and molecular biologists to work together, to devise new analytical methods with which to tease trustworthy signals from these data. My guess is that it will be quite a while before we can confidently determine the position of *Kenyanthropus platyops* in the human evolutionary tree. ■

Daniel E. Lieberman is in the Department of Anthropology, The George Washington University, 2110 G. Street NW, Washington DC 20052, USA.

e-mail: danlieb@gwu.edu

1. Wood, B. A. *Koobi Fora Research Project, Volume 4: Hominid Cranial Remains* (Clarendon, Oxford, 1991).
2. Strait, D. S., Grine, F. E. & Moniz, M. A. *J. Hum. Evol.* **32**, 17–82 (1997).
3. Leakey, M. G., Feibel, C. S., McDougall, I. & Walker, A. C. *Nature* **376**, 565–571 (1995).
4. Asfaw, B. *et al.* *Science* **284**, 629–635 (1999).
5. Brunet, M. *et al.* *Nature* **378**, 273–275 (1995).
6. White, T. D., Suwa, G. & Asfaw, B. *Nature* **371**, 306–312 (1994).
7. Senut, B. *et al.* *CR Acad. Sci.* **332**, 137–144 (2001).
8. Leakey, M. G. *et al.* *Nature* **410**, 433–440 (2001).
9. Wood, B. A. & Collard, M. C. *Science* **284**, 65–71 (1999).
10. Lieberman, D. E., Pilbeam, D. R. & Wood, B. A. *J. Hum. Evol.* **30**, 97–120 (1996).

Materials science

A pore view of corrosion

Martin Stratmann and Michael Rohwerder

The mechanism by which an alloy corrodes into a potentially useful porous sponge is understood qualitatively, but quantitative predictions of its final structure have been lacking. A model for this has now been proposed.

Alloys have an important role in modern life. Bronzes, stainless steels and alloys of noble (inert) metals, to name but a few, are all designed to have various properties, such as mechanical strength, ductility or resistance to corrosion. Pure metals are almost never used for technical applications because of their softness. Conversely, the corrosion of more reactive materials can be significantly reduced by adding even a small amount of a more inert component. For example, iron–chromium alloys containing as little as 13% chromium can survive long periods of exposure to salt water, whereas pure iron corrodes quickly. Alloys are also useful economically for materials based on noble metals such as platinum or gold: adding less noble components to noble metals used for jewellery, electrical contacts and dentistry significantly reduces their cost.

An interesting class of alloys is those that can be selectively depleted or 'dealloyed', because one component is more reactive (less noble) than the other. This process leaves behind an intricate nanoporous structure made almost entirely from the noble component. Although dealloying is quite common, and can generate useful nanoporous materials, the underlying physics of the process remains poorly understood.

On page 450 of this issue, Erlebacher *et al.*¹ present a model that predicts all of the characteristic features of dealloying, including the pore size of the final structures.

For dealloying to occur, the less inert component has to be selectively dissolved — either naturally in acid or more quickly in an electrochemical system. By using the alloy as one of the electrodes in an electrochemical cell, and then applying a voltage, ions will be stripped from the electrode and dissolve in the electrolyte. For selective dissolution to occur, the electrode potentials at which the alloy's two metals form ions must be significantly different, allowing one to dissolve in the electrolyte while the other remains intact. For example, gold–copper will dealloy, but gold–platinum will not.

Although studies of dealloying have been largely prompted by concerns about corrosion, the high surface area of dealloyed materials also makes them potentially useful as catalysts or sensors. The final structure usually has a spongy nature, consisting of a system of interconnected pores or tunnels in a skeleton of filaments of the pure, or almost pure, inert metal. Previous studies have shown that, depending on the alloy and electrolyte used, the pore and filament size is typically around 5–50 nm, and surface areas as high as 20 m² g^{−1} are possible. These

materials are similar to the naturally occurring, porous zeolites, which have pore sizes in the range of 1–2 nm and specific surface areas of some 100 m² g⁻¹. Despite this qualitative picture of dealloyed materials, a quantitative understanding of the mechanisms driving pore formation is missing. Erlebacher *et al.*¹ tackle this problem by developing a numerical model that can predict the pore and filament size of dealloyed structures, and explains several features of the electrolytic process.

Before discussing selective dissolution in detail, we need to introduce two concepts: the 'parting limit' and the 'critical potential'. The parting limit has a long history, dating back to work by Tammann² in 1919. According to Tammann, dealloying occurs if the less inert component exceeds a specific concentration limit, the parting limit. Seeking an explanation for this threshold, Masing³ calculated the probability with which a continuous filament of active (less inert) atoms would dissolve in a randomly orientated alloy in solution. As expected, this probability depends strongly on the concentration of the active atoms.

But such experiments revealed that the final nanoporous structure is not based on any pre-existing structures in the alloy. Somehow, the etching by the acid electrolyte creates a new network of tunnels and pores. Also, it is not sufficient for an alloy to contain a continuous filament of reactive atoms if the filament is too narrow to admit water or ions into the resulting channel. So, dealloying must involve some movement (diffusion) of the more noble component at the surface to open up channels for water ingress. Consequently, such dynamic processes are an important element of the numerical model developed by Erlebacher and colleagues.

Experiments also show that dealloying will not occur — even when an alloy greatly exceeds the parting limit — if the electrode potential is below a certain value, known as the critical potential, E_c . Below E_c the current of dissolved ions from the alloy is very low, but above E_c it rises rapidly and results in substantial dealloying as the structure is rapidly stripped of its less inert component. The value of E_c depends on the concentration of the inert component and increases as its concentration increases.

Previous models to explain dealloying addressed E_c within the framework of a percolation theory. One of its main features was the competition between the roughening effect on the surface brought about by selective depletion and the smoothing action of surface diffusion^{4,5}. Such models can predict E_c quite accurately, but give no information about the final size of the pores. Erlebacher *et al.*'s model¹ considers a minimum set of physical processes — the movement of atoms at the alloy surface and the movement of the dissolved ions in the electrolyte. This allows them to predict the concentration-dependent values of E_c and also the size of the pores in the dealloyed material.

There are still questions to be resolved. Dealloying usually refers to selective dissolution at potentials above E_c . But selective dissolution also occurs below E_c . No large-scale attack on the alloy is observed, but scanning tunnelling microscopy (STM) measurements reveal nanoscopic dissolution and diffusion at the surface^{6,7} (Fig. 1). The STM investigations show that in these early stages of pore formation the initial processes and structures are strongly dependent on the existing crystal structure and atomic order.

The work of Erlebacher *et al.* is an import-

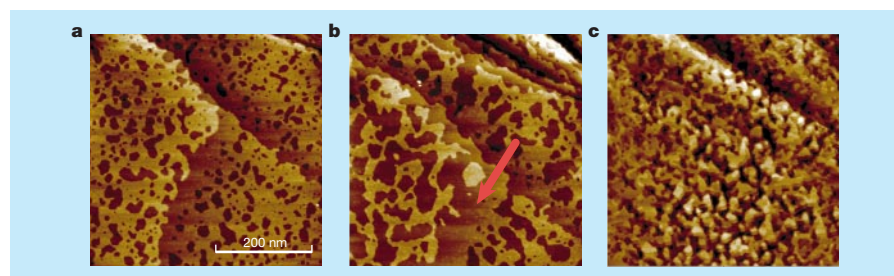


Figure 1 The early stages of dealloying of a copper–gold alloy immersed in a solution of sulphuric acid. In these STM images the darker patches are monatomic pits, whereas the light patches are predominantly gold atoms. Three monatomic terraces can be seen. At electrode potentials below the critical potential E_c the dealloying proceeds very slowly. a, Initially, two-dimensional dissolution nuclei are formed in the terraces, as the steps are already protected by gold atoms. b, Besides surface diffusion, vertical diffusion processes play an important role, as can be seen from the uncovered part of the second layer terrace that is not pitted (arrow). Here, the dissolution reaction is strongly inhibited because copper atoms have been replaced by gold from the receding terrace above. The question is how these processes lead to the three-dimensional nanoporous structures that develop at potentials above E_c . In their model, Erlebacher *et al.*¹ show that a higher potential rapidly strips the terraces of copper atoms, exposing all the surface gold atoms. The subsequent movement of these gold atoms exposes unprotected copper, which is etched away to form deep pits and eventually pores. c, This image was taken at a higher electrode potential (still well below E_c) in which the early formation of deeper pits can be seen.



100 YEARS AGO

The issue of the *Revue Scientifique (Revue Rose)* of March 9 contains a long and interesting article by M. Louis-Adrien Levat on the destruction of birds, especially by means of traps and snares, which he declares to be illicit ... during a single spring a few years ago no less than 1500 nests were taken in one French province. This represents a prospective loss of about 6000 birds, which might be expected to consume some 6,000,000 insects among them. He adds the significant observation that in the year 1860 one hundred cages filled with insectivorous birds of various kinds were exported from Baden to New South Wales; and that at the present day it would be almost impossible to send such another cargo, owing to the scarcity of these birds on the continent. And it is not alone the disappearance of bird-life and bird-song from the country districts that is to be deplored. The effects on agriculture, horticulture and the grape industry are simply disastrous. Some birds, it is computed, will consume 200,000 insects per season, and others as many as 600 per day... In Hérault alone the destruction of insectivorous birds is calculated to cost the department 100,000 hectolitres of wine annually.

From *Nature* 21 March 1901.

50 YEARS AGO

The 'membrane', that is, the electric double layer, surrounding the muscle fibre generates no field of forces inside the system. If, however, part of the membrane becomes depolarized, the depolarized part acts ... as a 'window' in the closed system and an electric field is generated. If the depolarization should occur with a sharp border, the maximal intensity of the field in the axis of the fibre would correspond to a field generated between the two plates of a condenser placed at a distance equal to the diameter of the fibre and having the same potential difference as the two sides of the 'membrane'. If the wave of depolarization travels along the membrane, electrodes along the axis charged correspondingly. Experiments, performed along this line by St. Hajdu and one of us, show that such a field will actually bring the contractile matter directly to contraction even if the membrane is inoperative... There seems to be thus no 'transmission' of excitation from the membrane to the contractile matter; depolarization and the elicitation of contraction merge into one single physical event.

From *Nature* 24 March 1951.

ant advance in predicting the sizes of pores in dealloyed structures, allowing the design and fabrication of nanoporous structures with adjustable pore sizes from 3 nm up to 5 μm . In biosensor applications, this may prove useful for immobilizing a wide range of biomolecules — such as the enzyme glucose oxidase — without the need for polymers such as polypyrrole. These polymers are typically used to immobilize enzymes on flat gold electrodes and are known to limit the lifetime of the sensor. ■

Martin Stratmann and Michael Rohwerder are at

the Max-Planck-Institut für Eisenforschung,
Postfach 140 444, D-40074 Düsseldorf, Germany.
e-mails: stratmann@mpie.de
rohwerder@mpie.de

1. Erlebacher, J., Aziz, M. J., Karma, A., Dimitrov, N. & Sieradzki, K. *Nature* **410**, 450–453 (2001).
2. Tammann, G. Z. *Anorg. Allg. Chem.* **107**, 1–239 (1919).
3. Masing, G. Z. *Anorg. Allg. Chem.* **118**, 293–308 (1921).
4. Sieradzki, K. J. *Electrochem. Soc.* **140**, 2868–2872 (1993).
5. Wagner, K. et al. *J. Electrochem. Soc.* **144**, 3545–3555 (1997).
6. Eckstein, G. A. In-situ rastertunnelmikroskopische Untersuchungen zur selektiven Korrosion von niedrigindizierten Au,Cu(hkl)- und Cu,Au(hkl)-Legierungseinkristallen. Thesis, Univ. Erlangen (2001).
7. Meyer-Streng, O. *Max-Planck Res.* **1**, 40–44 (2001).

Developmental biology

A twist on embryonic signalling

Richard M. Harland

Gradients of signalling proteins control the formation of different tissues during animal development. A controversial model for how one such gradient is produced now receives more experimental support.

During animal development, organizing centres at key locations in the embryo secrete proteins that dictate the pattern of cell types, tissues and organs. Such a protein may become distributed in a gradient of activity, and different concentrations instruct target cells to adopt different fates. The activity of such a 'morphogen' in a given embryonic region is determined by where and how much of it is made; its rate of diffusion or transport about the embryo; where and how quickly it is degraded; and the effects of any neutralizing molecules (antagonists). Writing on pages 475–487 of this issue^{1–3}, three groups take another step forwards in understanding the elaborate regulation of some potent morphogens — members of the bone morphogenetic protein (BMP) family.

The BMP developmental signalling pathway has been highly conserved in evolution, and is present, for example, in all the animals whose genomes have been sequenced — worms, fruitflies, mice and humans. In fruitfly (*Drosophila melanogaster*) embryos, the activity of the BMP molecule Decapentaplegic (Dpp) defines at least three different tissue types in the outer layer, or ectoderm⁴. Where Dpp signalling is absent, neuron-forming ectoderm can develop. At intermediate levels of Dpp, the ectoderm along the back forms. And peak levels of Dpp are needed for the formation of a membrane outside the embryo, the amnioserosa.

In the new papers, Scott *et al.*¹, Ross *et al.*² and Chang *et al.*³ look at the role of the Twisted gastrulation (Tsg) protein in both vertebrates and invertebrates, and show that this protein binds to members of the BMP family, preventing them from signalling. The story is intriguing because this Tsg activity was by no means obvious from analysis of flies that lack Tsg function. It was from such

studies that the *tsg* gene was first identified⁵. Normally, the process of 'gastrulation' in fruitflies leads to the three-layered structure of the larva, with the gut inside, the ectoderm outside, and mesoderm in between. Because of defects in the amnioserosa, this stage is abnormal in flies with mutations in *tsg* — hence the name of the protein.

Given that peak Dpp levels are needed for amnioserosa to form, and that loss of *tsg* causes loss of amnioserosa, it seemed plausible that the normal Tsg protein would somehow promote peak Dpp signalling. But it became apparent that this view might be too simple, because a Dpp-neutralizing molecule, Short Gastrulation (Sog), is also needed for amnioserosa development — and so, paradoxically, for peak Dpp activity⁴. Indeed, when Sog is experimentally expressed in a restricted region of the embryo, against a uniform background of the other proteins, a sharp peak of Dpp activity is produced at a distance from the Sog source⁶. How could this come about? The new papers^{1–3} help to provide a molecular explanation, which involves an interaction between Sog, Tsg and Dpp, and the breakdown of one component of this tripartite complex by a protein-degrading enzyme (Fig. 1).

The three groups looked at the involvement of Tsg in BMP signalling in *Drosophila*, *Xenopus laevis* (African clawed frog) and *Danio rerio* (zebrafish). In *Drosophila*, Tsg by itself can block the activation of a downstream signalling protein, Mad, by Dpp in cultured cells². And the interaction of Tsg with Sog converts Sog from a poor antagonist of Dpp to a highly effective one.

But why are both Sog and Tsg required for peak levels of Dpp signalling in the amnioserosa? It is assumed that Sog and Tsg

form a complex that helps Dpp to diffuse through the embryo, in part by preventing its association with cell-surface receptors along the way⁴. Tsg also accelerates the degradation of Sog by a protease (named Tolloid)^{1,7}. So, as suggested in Fig. 1, Tsg would accelerate the degradation of Sog throughout the embryo, releasing Dpp. But near the source of Sog synthesis, there would be so much Sog around that any released Dpp would be recaptured. At a distance from its source, degradation of Sog would restrict the amount of inhibitory complex, and Dpp would be dumped onto its cell-surface receptors. Although not intuitively obvious, this mechanism ensures that Dpp activity does not simply form a smooth gradient but instead reaches a sharp peak of activity away from the Sog source^{2,6}.

This mechanism differs from simple models for establishing gradients of morphogens, where a signal diffuses away from a

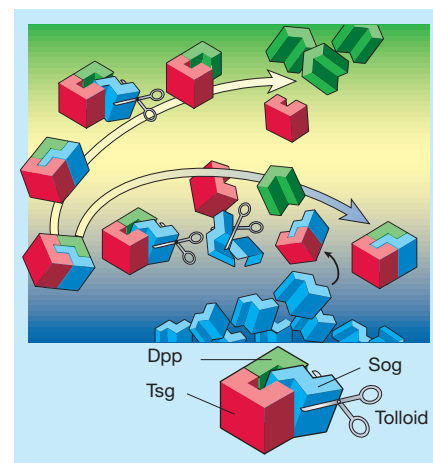


Figure 1 During an early stage in the development of the fruitfly *Drosophila melanogaster*, different ectodermal tissues are induced by high, intermediate and low levels of activity of a bone morphogenetic protein, Decapentaplegic (Dpp)⁴. This protein is produced through most of the embryo. Its levels are controlled by local production of Sog. Three new papers^{1–3} suggest how this control may be achieved. The Tsg protein and Dpp form a diffusible complex^{1–3,8}, as do Tsg and Sog^{1,2,8} (not shown). (These interactions also occur in vertebrates^{1–3,8}, where Chordin, a relative of Sog, also forms a complex with BMP4, a relative of Dpp¹⁰.) So, Tsg, Sog and Dpp can form a complex (left), in which Tsg and Sog inhibit Dpp. The protein-degrading enzyme Tolloid breaks down Sog, releasing Dpp in its active form^{11,12}; Tsg accelerates this process^{12,7}. In regions where Sog levels are high (bottom), degradation of Sog will have little effect, because there is plenty more around to form a complex with Tsg and Dpp. But, when Sog in the complex is degraded in areas where free Sog is scarce, Dpp will be released (top). The contribution of Screw, another bone morphogenetic protein, which is also blocked by Sog^{4,13}, is not shown.

local source, and is degraded at a local sink. But evolution does not start with models or design; instead it cobbles together components that happen to work. In flies, Dpp, Sog and Tsg together provide exquisite regulation of ectoderm development. Is the same true of vertebrates?

So far, the tools available to study the control of BMP signalling in vertebrates have not revealed all the details seen in flies. But the main principle — that Tsg helps the vertebrate equivalent of Sog (Chordin) to block BMP function — seems to be conserved. The three new papers^{1–3} reveal that, when injected into frog and zebrafish embryos, Tsg blocks BMP signalling. Moreover, when overexpressed in frogs, Tsg blocks the transcriptional targets of BMP signalling^{1,3}. Yet it does not induce an extra set of dorsal embryonic structures — the secondary axis that results from the absence of BMP signalling. This may in part be because of Tsg's diffusibility, as it can induce secondary axes when artificially tethered to cell surfaces³.

Overexpressed Tsg and Chordin synergize to block the expression of BMP target genes and to induce secondary axes^{1–3}. Oddly, though, these effects are lost when the stoichiometry of the two proteins is skewed by excess Tsg. This peculiarity remains to be understood. One possibility is that under some conditions, Tsg helps to displace BMPs from the fragments resulting from degradation of Chordin, and so promotes BMP signalling⁸, but this is controversial¹. Studies of vertebrates with mutations in *tsg* should

help to clarify the picture. Meanwhile, Ross *et al.*² have succeeded in eliminating Tsg function in zebrafish embryos by injecting antisense oligonucleotides, and it seems that Tsg is required, together with Chordin, to form the full complement of dorsal structures, supporting the conclusion that Tsg is a BMP antagonist.

Like most good developmental signalling molecules, Tsg is expressed throughout development^{1,3,8}, with some intriguing sites of expression in the skeleton¹. Although its function there has not yet been studied, proper skeleton formation is known to require fine-tuning by other BMP antagonists⁹. It will be interesting to see how they all interact.

Richard M. Harland is in the Department of Molecular and Cell Biology, University of California, 401 Barker Hall, Berkeley, California 94720-3204, USA.

e-mail: harland@socrates.berkeley.edu

1. Scott, I. C. *et al.* *Nature* **410**, 475–478 (2001).
2. Ross, J. J. *et al.* *Nature* **410**, 479–483 (2001).
3. Chang, C. *et al.* *Nature* **410**, 483–487 (2001).
4. Podos, S. D. & Ferguson, E. L. *Trends Genet.* **15**, 396–402 (1999).
5. Zusman, S. B. & Wieschaus, E. F. *Dev. Biol.* **111**, 359–371 (1985).
6. Ashe, H. L. & Levine, M. *Nature* **398**, 427–431 (1999).
7. Yu, K. *et al.* *Development* **127**, 2143–2154 (2000).
8. Oelgeschläger, M., Larrain, J., Geissert, D. & De Robertis, E. M. *Nature* **405**, 757–763 (2000).
9. Brunet, L. J., McMahon, J. A., McMahon, A. P. & Harland, R. M. *Science* **280**, 1455–1457 (1998).
10. Piccolo, S., Sasai, Y., Lu, B. & De Robertis, E. M. *Cell* **86**, 589–598 (1996).
11. Marques, G. *et al.* *Cell* **91**, 417–426 (1997).
12. Piccolo, S. *et al.* *Cell* **91**, 407–416 (1997).
13. Dorfman, R. & Shilo, B.-Z. *Development* **128**, 965–972 (2001).

charge of the electron, which results in a repulsive force between all of the electrons in the circuit. For example, for very small capacitors, adding a single electron to the many already inside the capacitor plates may be sufficient to raise the energy of the capacitor to prevent the next electron from joining. In general, the motion of any electron in a mesoscopic circuit can influence all other electrons in the circuit, making the calculations complicated.

Finally, we have to account for correlations in the motion of the electrons, which give rise to phenomena such as superconductivity.

The theory presented at the meeting by Nazarov can handle these complications, and should unite all the partial results obtained by theorists to date. But this does not mean that the work is finished. Nazarov uses the tools of quantum-field theory (used to describe elementary particles) to predict how electrons collectively cross mesoscopic circuits. However, like fundamental particle theory, the equations can be solved only in simple situations. More work will be needed to demonstrate its usefulness in predicting observable quantities in a circuit.

To take measurements on a mesoscopic system, wires are usually attached to the circuit to measure the electrical conductance. More recently, many researchers have realized that more information about the behaviour of electrons inside a circuit can be obtained by investigating the intrinsic noise in the current. For example, the noise resulting from fluctuations in the number of electrons traversing the circuit is proportional to the unit of charge in the circuit. The noise is like the sound of hailstones on the roof: the larger the stones, the greater the noise.

This idea has been confirmed by an experiment with 'Andreev reflection' — the doubling of the effective charge when an electron crosses over from a metal to a superconductor. In the experiment, a doubling of the noise was indeed observed (M. Sanquer, CEA Grenoble). For short junctions between two superconductors, multiple Andreev reflections were even observed to give rise to charge units of three, four and more times the electron charge (C. Urbina, CEA Saclay).

The noise studied to date measures only the lowest-order fluctuations in the current that arise by electrons crossing the circuit. However, much more information can be revealed by more detailed measurements. Measuring each individual electron that traverses the circuit, in the way that photons are counted in optics, should lead to a much better understanding of the interactions of many electrons inside the circuits. This may sound unrealistic. But the secret to success may be in counting photons, not electrons — an electron crossing over from one side of a circuit to the other can do so only by producing a low-energy photon. A recent

Mesoscopic physics

Noisy times ahead

Jan van Ruitenbeek

Realizing the dream of quantum computing will require knowledge and techniques from many fields. A new theory concerning the properties of nanometre-scale circuits can contribute.

Mesoscopic electronics is the study of the movement of electrons in tiny circuits — composed of conductors, capacitors and inductors — on the order of hundreds of nanometres to several micrometres. At these sizes, and at low temperatures, quantum mechanics dominates all of the properties of circuit elements, so the familiar laws used by engineers in designing and analysing electronic circuits are useless. At a conference earlier this year*, a complete mesoscopic theory was presented to replace these conventional laws (Y. Nazarov, Delft University of Technology). Although the theory is so complicated it can only be applied to simple systems, it indicates that

noise inherent in the electrical currents can reveal the properties of circuits, perhaps leading to quantum computers.

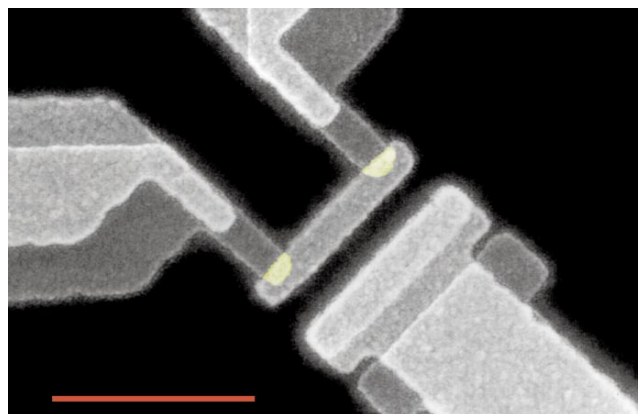
Over the past two decades, during which mesoscopic physics has emerged, theoretical work has provided an outline of how quantum circuits work. Properties of electrons cause three complications at mesoscopic scales. First, the motion of the electrons must be described in terms of quantum waves. Like light waves, quantum waves are scattered by objects such as the edge of the metallic conductor and defects or impurities present in its interior. This scattering results in an interference pattern, making properties such as resistance extremely sensitive to variations in the circuit geometry, even at the atomic scale.

The second complication arises from the

*Electronic Correlations: From Meso- to Nano-physics, XXXVI Rencontres de Moriond, Les Arcs, France, 20–27 January 2001. Abstracts are at www.lps-ipsud.fr/moriond01

Figure 1 Mesoscopic device for reading information in a qubit — a quantum bit.

This scanning electron micrograph shows a single electron transistor made from layers of aluminium and aluminium oxide that can be used as a read-out device for a qubit. The electrodes at the top left are the source and drain of the transistor. These electrodes connect the



central part of the transistor (the island) through tunnel junctions (the overlaps of the drain and source with the island; yellow). The size of each junction is about 50 nm by 60 nm. The island is capacitively coupled to an electrode (bottom right) that influences the current through the transistor. This type of device can be used as a fast, sensitive voltmeter with an extremely high input resistance. With optimized circuitry it could, in principle, detect about 1% of an electron in less than 1 microsecond. This should allow a qubit to be read out, without prematurely destroying it. (Red bar: 500 nm.) (Courtesy of the Quantronics Group, CEA Saclay, France.)

proposal¹ may bring the counting of these photons within reach.

In order to calculate subtle fluctuations in current, or the time distribution of low-energy photons, the analogy between electrons and light can be exploited further (C. W. J. Beenakker, Universiteit Leiden). Beenakker has derived an expression describing the relation between the properties of the circuit and the statistical distribution of the photon counts, so experiment and theory may now usefully be combined. Even the lowest-order fluctuations measured so far have already produced a wealth of information. Experimentalists are now considering attacking higher-order fluctuations, although serious technical challenges remain.

But where is all this leading us? One attractive application is quantum computing. The quest to create a quantum computer has spread into many areas of physics, including quantum optics and atomic physics. But the mesoscopic physics approach has the advantage that we can design and fabricate circuits consisting of many individual elements.

The quantum bits (qubits) of choice are superconducting circuits of capacitors or inductors containing weak superconducting links. A qubit must be formed from the superposition of two quantum states. For mesoscopic qubits, the quantum states are either flux states^{2,3} (where the current through a tiny ring can flow clockwise or anticlockwise) or charge states⁴ (where a capacitor can be either charged with one electron pair or uncharged).

The detection system needed to read out information in the qubits is critical (D. Esteve, CEA Saclay; R. Schoelkopf, Yale University; Fig. 1). The problem here is that the read-out circuit should be coupled to the quantum system, but this coupling unavoid-

ably perturbs the quantum state. Esteve and Schoelkopf show that the coupling can be made sufficiently weak for a reliable read-out, without premature destruction of the quantum state during the computation (Fig. 1). The next step will be to couple several

quantum bits and to show that we can create the desired 'entanglement' or interaction of the quantum states of the qubits, which is required for computation. Noise is the ideal tool to analyse this entanglement (D. Loss, University of Basel) because it reflects the statistical distribution of the electrons imposed by the quantum state of the circuit.

The challenge of developing a full-size quantum computer puts our ability to fabricate circuits and our understanding of these systems to an extreme test. It is becoming increasingly likely that to make a quantum computer we will have to integrate the knowledge and techniques from many fields, combining tools from mesoscopic physics with those from quantum optics and molecular physics.

Jan van Ruitenbeek is in the Faculty of Mathematics and Natural Science, Leiden Institute of Physics, Universiteit Leiden, NL 2300 RA Leiden, The Netherlands.

e-mail: ruitenbe@phys.leidenuniv.nl

1. Aguado, R. & Kouwenhoven, L. *Phys. Rev. Lett.* **84**, 1986–1989 (2000).
2. Nakamura, Y., Pashkin, Y. A. & Tsai, J. S. *Nature* **398**, 786–788 (1999).
3. van der Wal, C. H. *et al.* *Science* **290**, 773–777 (2000).
4. Bouchiat, V., Vion, D., Joyez, P., Esteve, D. & Devoret, M. *Phys. Scripta* **T76**, 165–170 (1998).

Immunology

Telling the brain about pain

Tamas Bartfai

Injury to one part of the body often results in hypersensitivity to pain somewhere else. Unusually, this process seems to be coordinated by the brain in a way that does not involve the transmission of nerve impulses.

When a tissue is injured, inflammation often occurs, resulting in general feelings of illness¹ — with symptoms such as fever, lethargy, anorexia, and general muscle and joint ache — and an oversensitivity of nearby uninjured tissue to pain. All this was thought to occur by a mechanism coordinated by the brain and involving the transmission of nerve impulses from the injured region, through the spinal cord, to the brain. But Samad and colleagues, writing on page 471 of this issue², suggest that nerve impulses are not involved, and that the general feelings of illness result from the production of inflammatory signals inside the brain. Their results, as well as those of Ek and co-workers (page 430)³, also offer hints as to how this might happen.

Most limbs are packed with sensory nerve endings and can detect touch, heat and acute pain — sensations that are transmitted rapidly and accurately to the brain by the nerves. Drugs that silence sensory nerves work well to relieve acute pain. But, as Samad *et al.*² show, when local inflamma-

tion has occurred, blocking nerve impulses produces little pain relief.

Samad and co-workers used a mild chemical to induce inflammation in the hind paws of rats. They found that this 'local' inflammation causes a rapid, large and long-lasting increase in the concentration of an inflammatory signalling molecule called interleukin-1 β in the cerebrospinal fluid, which fills the cavities in the brain and spinal cord. When the authors blocked interleukin-1 β in these rats — either by injecting drugs into the spine that inhibit the production of interleukin-1 β in cerebrospinal fluid, or by using a molecule that blocks the interleukin-1 β receptor on brain cells — they found that hypersensitivity to pain was strongly inhibited. This inhibition was less effective when the same compounds were injected directly into the bloodstream. In other words, it seems that it is mainly the increase in cerebrospinal levels of interleukin-1 β that tells the brain about local inflammation.

Traditionally, 'humoral' signalling molecules — such as interleukin-1 β produced at an injured tissue — are thought to be

released into the bloodstream, but not to travel into the brain because a layer of cells known as the blood–brain barrier prevents this. Samad *et al.*'s results leave open the question of how (and which) signals from the site of inflammation trigger the production of interleukin-1 β in cerebrospinal fluid. But the results of Ek *et al.*³ may help to provide an answer. The cells that make up the blood–brain barrier have on their surface receptors that specifically recognize interleukin-1 β . Ek *et al.* show that in response to interleukin-1 β — injected directly into the veins of rats — these cells express two enzymes, cyclooxygenase-2 (ref. 4) and prostaglandin E synthase. There is some discrepancy with Samad *et al.*'s results here, and the quantities of interleukin-1 β injected into the veins by both groups are much higher than those that would be produced during local inflammation. So the blood-borne inflammatory molecules involved remain unknown. But Ek *et al.*'s results do establish that cells of the blood–brain barrier have the capability to express cyclooxygenase-2 and prostaglandin E synthase during inflammation.

This means that, in response to inflammation, the cells of the blood–brain barrier have all three of the enzymes needed to turn phospholipids in their membranes into the potent inflammatory mediator prostaglandin E₂. The first enzyme needed is phospholipase A₂, which is present in most cell types and produces arachidonic acid from the phospholipids. Cyclooxygenase-2 then converts arachidonic acid into prostaglandin H₂. Finally, prostaglandin E synthase converts prostaglandin H₂ into prostaglandin E₂.

This last molecule is a highly active, inflammatory, fever-inducing and pain-causing signal^{5,6}. It is also soluble in lipids, and so, by passing through the outer membrane of the cells of the blood–brain barrier, it can enter the brain and cerebrospinal fluid. Once there, prostaglandin E₂ can activate its receptors on both neurons and microglia (immune cells in the brain), so 'exporting' the inflammation from the injured tissue to the brain. When prostaglandin E₂ binds to the EP1- and EP3-type prostaglandin receptors^{6,7} on neurons, it increases neuronal excitability, altering the threshold for the processing of sensory information such that non-painful stimuli become painful. Prostaglandin E₂ also activates the synthesis of interleukin-1 β by microglia, explaining how it can be found in the cerebrospinal fluid². The interleukin-1 β leads to production of cyclooxygenase-2 by neurons in the brain and spinal cord², and further synthesis of prostaglandin E₂, but now within the brain. This augments both the prostaglandin E₂ signal and the hypersensitivity to pain.

Cyclooxygenase-2 is the target of most antipyretic (fever-reducing) drugs, as well as

many new anti-inflammatory drugs — two of which, VIOXX and Celebrex, have seen some success in reducing the pain resulting from arthritis, for example. When developing this class of painkillers, researchers only intended them to work at locations outside the brain, so made no special effort to ensure that the drugs easily penetrate the blood–brain barrier. But, as Samad *et al.* show², if cyclooxygenase-2 both inside and outside the brain is inhibited, then better pain relief is achieved. Their results suggest that inhibitors of cyclooxygenase-2 that better penetrate the blood–brain barrier would be even more efficient painkillers,

working to block the oversensitivity to pain that is coordinated by the brain. ■

Tamas Bartfai is in the Department of Neuropharmacology, Neurological Research Center, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, USA. e-mail: tbartfai@scripps.edu

1. Dantzer, R. *et al.* *Ann. NY Acad. Sci.* **856**, 132–138 (1988).
2. Samad, T. A. *et al.* *Nature* **410**, 471–475 (2001).
3. Ek, M. *et al.* *Nature* **410**, 430–431 (2001).
4. Vane, J. R., Bakhle, Y. S. & Botting, R. M. *Annu. Rev. Pharmacol. Toxicol.* **38**, 97–120 (1998).
5. Griffiths, R. in *Inflammation: Basic Principles and Clinical Correlates* (eds Gallin, J. & Snyderman, R.) 349–360 (Lippincott Williams and Wilkins, Philadelphia, 1999).
6. Stock, J. L. *et al.* *J. Clin. Invest.* **107**, 325–331 (2001).
7. Ushikubi, F. *et al.* *Nature* **395**, 281–284 (1998).

Glaciology

Enigmatic Arctic ice sheets

Robert Spielhagen

To what extent was the Arctic Ocean glaciated in the past? Heavily, according to data, gathered by a submarine, which show considerable ice-scouring of topography in parts of the ocean basin.

Those of us living in the northern parts of Europe or North America can see the legacy of the ice ages from a brief look out of the window. Several times during the past million years, huge ice domes, kilometres thick, have covered vast areas of the Northern Hemisphere. On land, U-shaped mountain valleys, moraines and dislocated rocks are obvious signs of ice-sheet activity in the past. But what happened when the ice sheets reached the sea?

Most glaciologists believe that the ice-sheet fringes rapidly disintegrated, releasing icebergs into the adjacent ocean. But on page 453 of this issue¹, Polyak *et al.* report evidence to the contrary. They describe features in the depths of the Arctic Ocean, discovered from a submarine by acoustic techniques, that are evidence of erosion caused by ice sheets more than 1,000 kilometres off the present-day continental coastlines. The authors conclude that during some glaciations the ice sheets in northwestern Eurasia and the Canadian islands covered a very large area, encroaching well over the Arctic Ocean. Such ice sheets may have been like the wide, partly floating ice-shelves that exist off Antarctica today. But this news comes at a time when results of field work in arctic Russia and Alaska^{2,3} are suggesting that the northern continental ice sheets were of only limited extent (Fig. 1), or even absent, during the Last Glacial Maximum (LGM), which occurred around 20,000 years ago.

The broader context here is that for some years Grosswald and Hughes^{4,5} have proposed that a huge 'Eurasian ice sheet' existed during the LGM. This ice sheet is proposed to have connected the ice domes covering Scandinavia and the British Isles with those

over Alaska, as shown in Fig. 1, and included floating ice shelves over the Arctic Ocean. In reduced form, this Eurasian ice sheet was incorporated into many palaeoclimatic and palaeotopographic simulations of the LGM^{6,7}. However, many of the glacial features in northern Siberia have turned out to be much older. They are now ascribed to

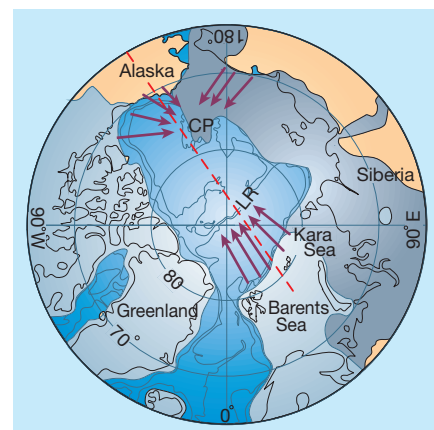


Figure 1 Possible ice-sheet distribution around the Arctic Ocean during the Last Glacial Maximum, about 20,000 years ago. This reconstruction is based on refs 2 and 11 (ice coverage in light grey) and ref. 4 (ice coverage in dark grey). Blue indicates sea-ice-covered or open water. Polyak and colleagues¹ propose that ice coverage was much greater than shown here, but at times that have not yet been specified. The proposed movement of the floating ice shelves that constituted the greater coverage is indicated by arrows. The broken red line indicates the profile displayed in Fig. 2, overleaf. The thicker black lines indicate coastlines, the thinner ones bathymetry (500- and 2,000-m isobaths). CP, Chukchi Plateau; LR, Lomonosov Ridge.

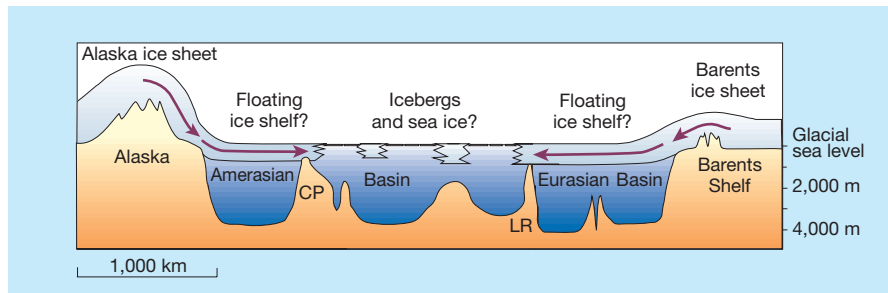


Figure 2 Profile across the Arctic Ocean, as indicated by the red line in Fig. 1. The graphic indicates the topography of the ocean basin, and its subdivisions, and shows the possible extent of floating ice shelves inferred by Polyak *et al.*¹. These ice shelves may have developed asynchronously. CP, Chukchi Plateau; LR, Lomonosov Ridge.

earlier phases of the last interglacial–glacial cycle, which started 115,000 years ago, or to even older ice ages^{2,3}.

At first sight, the results of Polyak and co-workers¹ support the Grosswald–Hughes model. But stratigraphic models for sediments of the central Arctic Ocean are still controversial, so Polyak *et al.* are cautious in assigning ages to the newly found glacial features in the deep Arctic Ocean. Further sediment analyses will be necessary to address this uncertainty.

That apart, the new findings raise several questions that will have to be tackled by the combined expertise of glaciologists, geologists, climatologists and modellers. Can large ice shelves remain stable in an open ocean basin? What could have caused the rapid expansion of floating ice shelves across the deep, wide Arctic basins? Why were the sizes and configurations of the ice sheets in the Northern Hemisphere apparently so different in different cold periods? And what were the feedback mechanisms between ice-sheet development, ocean currents and atmospheric circulation?

There are two principal requirements for the birth and growth of glacial ice sheets: low temperatures, and circumstances in which snow precipitation exceeds the rate of melting in the summer. A reduction in the amount of solar energy reaching Earth set the stage for high-latitude cooling during ice ages, but it was moisture supply that was the decisive factor in determining where and when ice sheets formed. For example, a correlation between LGM ice-sheet growth in the northwestern Barents Sea and seasonally ice-free conditions in the adjacent Norwegian Sea has been documented⁸. Such findings indicate the key role of the oceans in ice-sheet growth.

According to Polyak *et al.*¹, an earlier ice sheet in the Barents and Kara seas produced a floating ice shelf, about 1 km thick, which reached and scoured the underwater Lomonosov Ridge in the central Arctic (see Fig. 2). At the same water depth (970 m), but further south, the ridge seems to have been unaffected⁹—so the ice shelf must have been tongue-shaped and thinner there, or must

have been bounded by sea ice to the east. Ice shelves, however, are unstable and are highly sensitive to variations in sea level. So the flow from the ice-supply routes of the continental shelves must have been extremely fast, which implies a much stronger moisture supply than during the LGM (when the Barents–Kara ice sheet terminated at the shelf break).

Then, as now, the most important moisture source for Eurasia was the North Atlantic. During the LGM, the North Atlantic was cold⁶—probably too cold to feed fast-flowing ice sheets. It could have been that conditions intermediate between glacials and interglacials (interstadials), when water temperatures and evaporation were higher than in full glacials, were more favourable to the growth of large ice sheets in the Eurasian Arctic. Indeed, during the interstadial of around 70,000 years ago, ice sheets reached much further to the east than at any time since then².

The new results from fieldwork on land and deep-sea surveys seem puzzling, but are not necessarily contradictory. The timing of pre-LGM glacial events remains poorly known, but it holds the key to understanding the behaviour and impact of ice sheets. At a time when many palaeoceanographers are investigating in great detail the influence of past ice sheets on ocean circulation and climate¹⁰, we have good reason to reconsider events in the Arctic: climatologists and modellers have much to learn about why, when and how large ice sheets formed there. ■

Robert Spielhagen is at GEOMAR, Research Center for Marine Geosciences, Wischhofstrasse 1–3, D-24148 Kiel, Germany.

e-mail: rspielhagen@geomar.de

1. Polyak, L., Edwards, M. H., Coakley, B. J. & Jakobsson, M. *Nature* **410**, 453–457 (2001).
2. Svendsen, J. I. *et al.* *Boreas* **28**, 234–242 (1999).
3. Brigham-Grette, J. *Quat. Sci. Rev.* **20**, 15–24 (2001).
4. Hughes, T., Denton, G. H. & Grosswald, M. G. *Nature* **266**, 596–602 (1977).
5. Grosswald, M. G. & Hughes, T. J. *Polar Geogr.* **23**, 23–54 (1999).
6. CLIMAP Project Members *Geol. Soc. Am. Map and Chart Ser. MC-36* (1981).
7. Peltier, W. R. *Science* **265**, 195–201 (1994).
8. Hebbeln, D., Dokken, T., Andersen, E. S., Hald, M. & Elverhøi, A. *Nature* **370**, 357–360 (1994).
9. Jokat, W. (ed.) *Rep. Polar Res.* **308** (1999).
10. Broecker, W. S. *Nature* **372**, 421–424 (1994).
11. Prest, V. K. *Geol. Surv. Can. Pap.* **84-10**, 21–38, map 1584A (1984).

Daedalus

Computing for art

Computer optimization is a powerful art form these days. Starting from a crude approximation in parameter space, an iterative program can ‘hill-climb’ very rapidly to a useful optimum. Daedalus now wants to apply it to the greatest art form of all, which these days is advertising. Enormous sums of money are cheerfully spent seeking the optimum advertisement that will grip potential purchasers, and the speed and lack of conventional wisdom of the computer is bound to help.

When Standard Oil changed its name to Esso, only human decision was needed. When the latter changed further to Exxon, a computer program churned out thousands of possible names, ignorant of the limitations of the letter ‘x’ in English, and a human observer reacted to the outcome. So, says Daedalus, what we need is a big screen, or many little screens, all watched by potential customers wired up to headphones, electroencephalograph, heartbeat and galvanic skin-response detectors, and so on, all of which are fed back to a master computer controlling the whole thing.

The initial screen image, maybe a boot or a bicycle, will be so faint that the subjects will probably take it to be an imagining of their own, if a bit self-willed. But, far faster than any focus group, the machine will take everybody through a whole range of options until its hill-climbing optimization routine has identified that sudden frisson of delight or recognition in at least one subject. The whole languages, the gigabytes of visual data, now easily stored and accessed in modern computers, make feasible the entire ambitious concept—indeed, much of the expense of the project will lie in setting up the initial database of resources. Fortunately these can be almost continuously upgraded with more words, music and even existing advertisements, as the project proceeds.

Thereafter, the machine will improve the sound and image to a powerful democratic optimum. The project will be far more expensive than the conventional activities of advertising agencies. But it will also be far faster and far more likely to produce great and powerful ads.

For his hill-climbing project, Daedalus has settled on advertising for its dominance and financial rewards. But his machine should also be able to optimize pop music, high art, architecture, and other art forms that also fall short of what one might reasonably expect. **David Jones**

Carbon sink for a century

Intact rainforests have a long-term storage capacity.

With fossil-fuel combustion and land-use activities threatening to double atmospheric carbon dioxide this century¹, maintaining large forests as carbon reservoirs becomes an additional conservation incentive. We have developed a stochastic-empirical model that simulates forest-carbon cycling and now use this model to explore the response of the central Amazonian forest to an increase in biomass productivity. Our results show that these trees will accumulate carbon in their wood for more than a century after a productivity increase, underscoring the value of intact tropical forests.

Evidence from eddy covariance studies², analysis of forest inventory plots³, and atmospheric inversion models⁴ indicate that undisturbed neotropical forests remove a significant portion of human-derived CO₂ emissions from the atmosphere. Although these results do not provide insight into causal factors, open-top chamber experiments indicate that CO₂ doubling enhances the production of woody tissue per unit leaf area by about 25% (ref. 5), and plant physiological ecology provides a mechanistic basis for this response⁶.

The response of forest carbon storage to higher productivity is a function of the fraction of production allocated to, and the residence time of, the component reservoir⁶. Large wood (trunks and branches of or over 10 cm diameter and roots over 5 cm diameter) in central Amazon forest represents about 45% of total carbon storage and 30% of above-ground production, with a mean residence time of 80 years. Thus, compared with other ecosystem components, large wood has a high storage capacity, with a relatively long lag time after disturbance. Passive soil organic matter also has a large storage capacity, but carbon movement into this pool is very slow⁷.

Our model⁸ was parameterized using extensive field data of individual trees larger than 10 cm base diameter (D_b), including recruitment, growth and mortality ($n=18$ hectares), tree allometry (327 trees)⁹, maximum size of trees ($n=220$ species), wood density ($n=268$ species), and wood decomposition¹⁰ rates (155 dead trees). The model starts by filling 20 × 20 m cells (stands) with a variable number of trees whose characteristics are assigned using random deviates from the empirical distributions. Each year, variation in recruitment, growth, mortality and decomposition determines the carbon balance of the entire plot (and aggregate of stands).

We demonstrated model reliability by

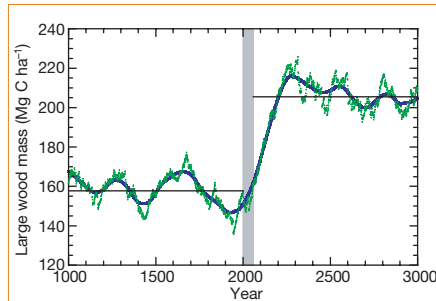


Figure 1 Example of large-wood carbon storage response predicted by the model⁸ to a 25% increase in biomass productivity over 50 years (0.25% per year). Cubic-spline fit shows averaged response; mean carbon storage (thin horizontal lines) stabilizes 127 ± 28 (95% C.I.) years (mean of eight 100-ha runs) after a 50-yr production increase (grey bar) with an accumulation rate of $0.5 \text{ Mg C ha}^{-1} \text{ yr}^{-1}$. Stability is defined as when any year's mass falls within a 95% C.I. of the mean (final 500 years) large-wood mass.

comparing predictions with the same field data (respectively), including large wood (predicted, 164 Mg C ha^{-1} ; field value, 156 Mg C ha^{-1}) and coarse litter (16 vs 15 Mg C ha^{-1}) carbon stocks, large wood growth (1.6 vs $1.7 \text{ Mg C ha}^{-1} \text{ yr}^{-1}$) and mortality (1.8 vs $2.1 \text{ Mg C ha}^{-1} \text{ yr}^{-1}$) rates, mean D_b (20.4 vs 21.1 cm), mean age for trees with greater than $100 \text{ cm } D_b$ (383 vs 425 yr), and maximum tree age¹¹ ($1,192$ vs $1,372 \text{ yr}$).

We used this model to investigate the response of large-wood carbon storage to a 25% increase in production distributed over 50 years for a 100-ha plot. As expected, carbon accumulated during years of increasing productivity. But once the productivity increase stopped (as a result of other limiting factors), large wood continued to accumulate carbon for over a century (Fig. 1).

Central Amazon upland (non-flooded environments) trees grow very slowly (mean, $1.1 \text{ mm } D_b \text{ yr}^{-1}$), and carbon storage only reached a new dynamic equilibrium once most trees had grown for their entire lives (mean age, 175 years) under conditions of enhanced productivity and the mass of a tree of mean D_b had stabilized at a higher value.

Forests respond in complex ways to increased atmospheric CO₂ and other global changes¹², and our model does not include potential limitations to a productivity increase or changes in vegetation characteristics. But for a given increase in productivity, it reveals that large-wood carbon storage will exhibit a lengthy lag time.

Avoiding deforestation will help industrialized countries to meet their Kyoto protocol emission-reduction obligations. Deforestation not only transfers carbon stocks directly to the atmosphere by combus-



Figure 2 The central Amazonian rainforest. The conservation value of tropical forests includes providing habitat for about half of the world's biological diversity, maintaining existing hydrological cycles and providing chemical precursors for the biomedical industry, not to mention the aesthetic and intrinsic value of a vast natural wilderness. The trees can also act as an important sink for carbon, continuing to accumulate it in their wood for over a century after an increase in biomass productivity.

tion, but it also destroys a valuable mechanism for controlling atmospheric CO₂.

Given a large increase in productivity, we calculate the rate of carbon sequestration to be of the order of 0.2–0.3 petagrams of carbon per year (which amounts to US\$2–3 billion per year at \$10 per Mg C) for large wood of the intact forest in Amazonia (Fig. 2), compared to fossil-fuel emissions of roughly 6 Pg of C per year. In all scenarios that reduce emissions to levels commensurate with the Kyoto protocol, numerous net biospheric exchanges of carbon are important. But terrestrial sequestration can only offer a partial solution, so controlling atmospheric CO₂ will require substantial reductions in fossil-fuel emissions in the coming decades.

Jeffrey Q. Chambers*†‡, Niro Higuchi§, Edgard S. Tribuzy§, Susan E. Trumbore*

*University of California Earth System Science, Irvine, California 92697-3100, USA

e-mail: chambersjq@yahoo.com

†National Center for Ecological Analysis and Synthesis, Santa Barbara, California 93101, USA

‡Biological Dynamics of Forest Fragments Project/INPA, Instituto Nacional de Pesquisas da Amazônia, Manaus, Amazonas 69011-970, Brazil

1. Tans, P. P. *et al.* *Science* **247**, 1431–1438 (1990).
2. Malhi, Y. & Grace, J. *Trends Ecol. Evol.* **15**, 332–337 (2000).
3. Phillips, O. L. *et al.* *Science* **282**, 439–442 (1998).
4. Keeling, R. F. *et al.* *Nature* **381**, 218–221 (1996).
5. Norby, R. J. *Plant Cell Environ.* **22**, 683–714 (1999).
6. Lloyd, J. *Funct. Ecol.* **13**, 439–459 (1999).
7. Trumbore, S. E. *et al.* *Glob. Biogeochem. Cycl.* **9**, 515–528 (1995).
8. Chambers, J. Q. *The Role of Large Wood in the Carbon Cycle of Central Amazon Rain Forest*. Thesis, Univ. California (1998).
9. Chambers, J. Q., Santos, J., Ribeiro, R. J. & Higuchi, N. *For. Ecol. Manage.* (in the press).
10. Chambers, J. Q. *et al.* *Oecologia* **122**, 380–388 (2000).
11. Chambers, J. Q., Higuchi, N. & Schimel, D. P. *Nature* **391**, 135–136 (1998).
12. Cox, P. M. *et al.* *Nature* **408**, 184–187 (2000).

Asexual reproduction

'Midwives' assist dividing amoebae

Asexual cells are normally able to reproduce entirely by themselves. But we have discovered that in about one-third of the dividing cells of *Entamoeba invadens* contraction of the cleavage furrow¹ may stop before separation is complete. We show here that the connected daughter cells overcome this problem by calling upon a neighbouring amoeba to help them achieve the final stage of division. The 'midwife' cell is chemotactically recruited for this mechanical intervention in what is a surprising example of primitive cooperation.

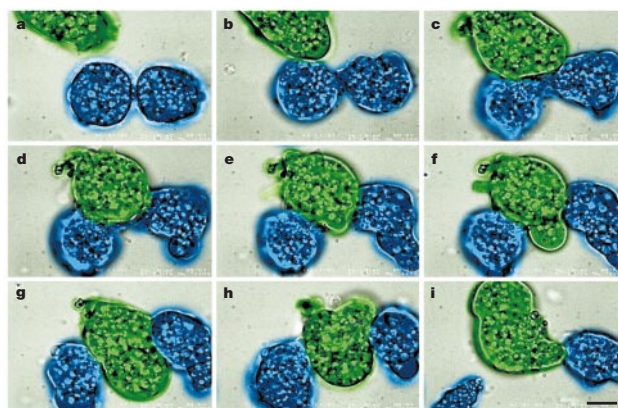
When an amoeba divides, the two daughter cells stay attached by a tubular tether which remains intact unless mechanically severed. If called upon, the neighbouring amoeba midwife (green cell in Fig. 1) travels up to 200 μm towards the dividing amoeba (see movie in supplementary information), usually advancing in a straight trajectory with an average velocity of $0.54 \pm 0.08 \mu\text{m s}^{-1}$. The midwife then proceeds to rupture the connection, after which all three amoebae move on.

Using four different samples, each at amoeba densities of about $3 \times 10^4 \text{ cells cm}^{-2}$, we recorded 106 divisions, of which 32 (30 $\pm$ 7%) were involved in cooperative reproduction (statistics were averaged over samples to give a 95% confidence level). In 11 cases (10 $\pm$ 5%), the division was aborted, and the amoebae continued to live as multinucleated cells. The multinuclear cells may resume their attempt to divide, sometimes producing three or four viable daughter cells^{2,3}.

In most cases (63 divisions, or 60 $\pm$ 8%), however, separation eventually occurred⁴, with both daughter cells resuming motility and using traction force to move away from each other, tugging at their connection so that it lengthened and narrowed until it finally broke. The incidence of abortions and bipolar divisions rises significantly (20 $\pm$ 5% aborted of 82 divisions) at low cell densities, when midwifing is rare.

To test whether the midwife is attracted chemotactically to the labouring division site, we aspirated fluid from the vicinity of the cleavage furrow of a dividing amoeba with a micropipette, then carefully discharged it near a distant amoeba. Of 41 such experiments, 15 (37%) revealed a positive chemotactic response (see movie in supplementary information; amoebae extended a directed pseudopod and followed a retracting pipette for a distance of over three times their own length); in 12 (30%), no response was seen; and in the remainder, cells moved in the general direction of the stimulus but resumed random

Figure 1 A typical 'midwifing' event, with the participants coloured for clarification. Frame times (a–i) are 0, 13, 30, 36, 41, 43, 49, 59 and 84 s, respectively. **a–c**, The midwife (green) aligns itself against the cleavage furrow; **d–f**, it puts out a pseudopod perpendicular to the furrow; **g–i**, then forcibly moves into the gap to disconnect the two daughter cells (blue). Although a few midwife amoebae may respond to the chemoattractant, only one, usually the nearest neighbour, is involved in completing the division. The midwife begins to move towards the dividing cell 2–3 min after division begins; this timing is compatible with an attractant with a diffusion coefficient of 10^{-6} – $10^{-7} \text{ cm}^2 \text{ s}^{-1}$ diffusing over a distance of 100 μm . Scale bar, 10 μm .



motion before traversing 100 μm .

Amoebae showing positive chemotactic responses pursued the pipette over distances of a few hundred micrometres, persistently following the tip when we changed the direction of the pipette's trajectory. Controls consisting of fluid aspirated from the vicinity of a non-dividing amoeba or of fresh medium alone elicited no response from neighbouring cells.

Identifying the chemical attractant in a volume of only 10 picolitres is difficult, so we sampled a mature culture of cells at the peak of their division cycle, reasoning that the concentration of chemoattractant would be higher than in a young culture. We found that fluid drawn from a confluent culture (3–4 days old, undergoing over a million divisions per flask per day) and transferred into a culture of only a few hours old, induced a strong chemotactic response in the younger cells (see movie). These amoebae followed retracting pipettes for over 30 min at an average linear velocity of $0.73 \pm 0.09 \mu\text{m s}^{-1}$ and flocked towards a stationary pipette at an average radial velocity of $0.34 \pm 0.04 \mu\text{m s}^{-1}$. Control solutions of cell-free three-day-old medium elicited no response; to control for any possible effect of starvation, fresh medium was changed 18 hours before collecting the fluid.

As we sampled fluid from cultures before the cells started clustering, we consider it most likely that the factor responsible for attracting midwives is the one secreted by cytokinesing cells, although it could be an entirely different chemoattractant. We have established that its relative molecular mass is between 50K and 100K and that it is still active after heating for 20 min at 95 $^{\circ}\text{C}$, but it is sensitive to oxidation by 0.24 M metaperiodate. Glycoconjugates, particularly membrane lipophosphoglycan-like molecules⁵, have these properties and are therefore likely candidates for the midwife-recruiting chemoattractant molecule.

A similar phenomenon has recently been observed in dividing cells of the amoeba *Dictyostelium discoideum* (G. Gerisch

and J. Köhler, personal communication). Midwifing may therefore constitute a more general adaptation of existing chemotactic mechanisms^{6,7}, enabling flaws in reproduction to be overcome through cooperation.

David Biron*, **Pazit Libros†**, **Dror Sagi***, **David Mirelman†**, **Elisha Moses***

Departments of *Physics of Complex Systems, and †Biological Chemistry, The Weizmann Institute of Science, Rehovot 76100, Israel

e-mail: elisha.moses@weizmann.ac.il

1. Schroeder, T. E. *Ann. NY Acad. Sci.* **582**, 78–87 (1990).
2. Weber, I. et al. *Curr. Biol.* **10**, 501–506 (2000).
3. Makioka, A., Kumagai, M., Ohtomo, H., Kobayashi, S. & Takeuchi, T. *Parasitol. Res.* **86**, 599–602 (2000).
4. Fukui, Y. *Microscopy Res. Tech.* **49**, 183–189 (2000).
5. Moody-Haupt, S., Patterson, J. H., Mirelman, D. & McConville, M. J. *J. Mol. Biol.* **297**, 409–420 (2000).
6. Bailey, G. B., Leitch, G. J. & Day, D. B. *J. Protozool.* **32**, 341–346 (1985).
7. Mato, T. M., Losada, A., Nanjundiah, V. & Konijn, T. M. *Proc. Natl Acad. Sci. USA* **72**, 4991–4993 (1975).

Supplementary information is available on Nature's website at <http://www.nature.com>.

Inflammatory response

Pathway across the blood–brain barrier

Inflammatory reactions against invaders in the body call upon cytokine molecules that elicit systemic responses, such as fever, fatigue, increased pain sensitivity and appetite loss, mediated by the central nervous system. But how cytokines can induce these effects has been a mystery as they are unlikely to cross the blood–brain barrier^{1–3}. Here we show that cerebral vascular cells express components enabling a blood-borne cytokine to stimulate the production of prostaglandin E_2 , an inflammatory mediator whose small size and lipophilic properties allow it to diffuse into the brain parenchyma. As receptors for this prostaglandin are found on responsive deep neural structures^{4–6}, we propose that the activated immune system controls central reactions to peripheral inflammation through a prostaglandin-dependent, cytokine-mediated pathway.

The final step in the biosynthesis of

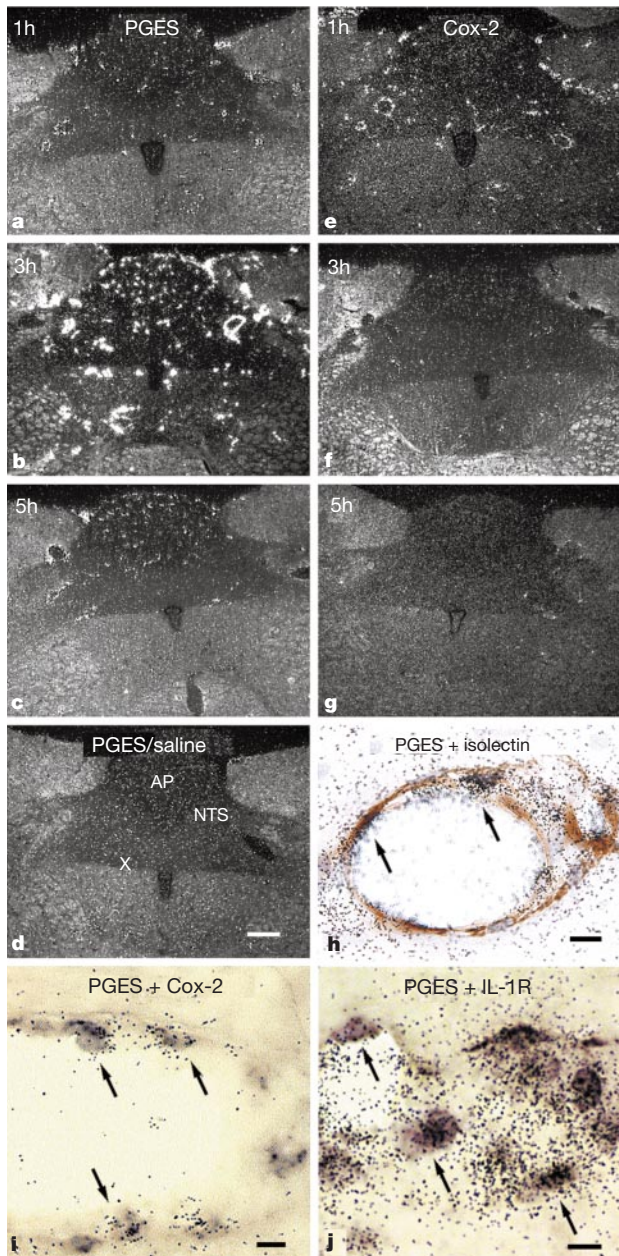


Figure 1 Kinetics and localization of cytokine-induced microsomal prostaglandin synthase (PGES), determined by *in situ* hybridization. **a–c**, Expression of microsomal PGES messenger RNA and **e–g**, of cyclooxygenase-2 (Cox-2) messenger RNA in the lower brainstem of the rat following intravenous injection of interleukin-1 β (2 μ g per kg body weight). The expression of microsomal PGES messenger RNA peaks at 3 hours after injection, whereas Cox-2 messenger RNA expression is maximal at 1 hour and almost undetectable later on. **d**, No microsomal PGES messenger RNA was induced in control animals injected with saline (at 3 hours post-injection). Method specificity was confirmed by the absence of radiolabelling after hybridization with microsomal PGES and Cox-2 sense-RNA probes (data not shown). AP, area postrema; NTS, nucleus of the solitary tract; X, dorsal motor nucleus of the vagus nerve. Scale bar, 200 μ m. **h**, Doubly labelled endothelial-like cells (arrows) stained for GSA B4 isoelectin (brown) and microsome PGES messenger RNA (silver grains). **i, j**, Co-localization (arrows) of microsome PGES messenger RNA (brown) with **i**, Cox-2 messenger RNA, or **j**, type-1 interleukin-1 receptor (IL-1R) messenger RNA (silver grains), respectively. Scale bars, 15 μ m.

prostaglandin E_2 involves the enzyme prostaglandin synthase⁷. We cloned the microsomal form of this synthase (mPGES) from rat, expressed it in a bacterial system, and confirmed that the protein was of the expected size (16K–17K) by western blotting and that it was fully functional in an enzyme assay.

We used *in situ* hybridization histochemistry to investigate the temporal and spatial expression of the mPGES gene in rat brain after intravenous injection of the pro-inflammatory cytokine interleukin-1 β . Expression of mPGES messenger RNA was detected in the cerebral blood vessels after one hour, peaked at 3 hours, and by 5 hours had returned almost to background levels (Fig. 1a–d). The kinetics of induction of mPGES expression in cerebral blood vessels thus matches the time course for biosynthesis of prostaglandin E_2 in the cerebral vas-

culature following immune stimulation⁸.

This induction of mPGES was accompanied by a rapid but more transient upregulation of mRNA encoding cyclooxygenase-2 (the inducible form of the enzyme catalysing the first step in the biosynthesis pathway), which peaked one hour after injection of interleukin-1 (Fig. 1e–g). This temporal difference between the induction of mPGES mRNA and cyclooxygenase-2 mRNA expression fits with the order in which these two enzymes are engaged in prostaglandin E_2 synthesis.

We characterized the cells expressing interleukin-1-induced mPGES mRNA by combining these results from *in situ* hybridization with radioactively labelled mPGES ('antisense') RNA probes with immunohistochemical staining with GSA B4 isoelectin, which labels endothelial cells, microglia and perivascular macrophages.

Doubly labelled cells were abundant, with most being flat like endothelial cells (Fig. 1h) and some being more rounded like perivascular macrophages.

As cyclooxygenase-2 is associated with mPGES *in vitro*^{9,10}, we tested whether these enzymes co-localized inside the same cell by using a dual *in situ* hybridization technique that simultaneously detects radioactively labelled and non-radiolabelled antisense-RNA probes. We found that almost all vascular cells expressing mRNA for cyclooxygenase-2 also showed an interleukin-1-induced upregulation of mPGES mRNA (Fig. 1i), indicating that these cells have a coupled system for producing prostaglandin E_2 from arachidonic acid. Most of the cells expressing mPGES mRNA also co-expressed mRNA encoding the type-1 receptor for interleukin-1 (Fig. 1j).

We conclude that circulating interleukin-1 induces the synthesis of prostaglandin E_2 by signalling through the cell-surface interleukin-1 receptor to the interior of endothelial cells and possibly of perivascular macrophages in the brain vasculature. Our results indicate that the distinct distribution of different receptor subtypes for prostaglandin E_2 within the brain parenchyma^{4–6} may account for the specificity of the cellular responses triggered by interleukin-1 in discrete brain structures¹¹. Microsomal prostaglandin synthase could present a new target for developing drugs that will not only treat inflammatory conditions and their accompanying acute-phase reactions controlled by the central nervous system, but which might also be free of the systemic side effects associated with cyclooxygenase inhibitors.

Monica Ek*, **David Engblom†**, **Sipra Saha‡**, **Anders Blomqvist†**, **Per-Johan Jakobsson‡**, **Anders Ericsson-Dahlstrand***

*Department of Medicine, Unit of Rheumatology, ‡Department of Medical Biochemistry and Biophysics, The Karolinska Institute, Stockholm 17177, Sweden

†Department of Biomedicine and Surgery, Division of Cell Biology, University of Linköping, 58185 Linköping, Sweden
e-mail: andbl@mcb.liu.se

1. Rothwell, N. J. & Hopkins, S. J. *Trends Neurosci.* **18**, 130–136 (1995).
2. Elmquist, J. K., Scammell, T. E. & Saper, C. B. *Trends Neurosci.* **20**, 565–570 (1997).
3. Ericsson, A., Arias, C. & Sawchenko, P. E. *J. Neurosci.* **17**, 7166–7179 (1997).
4. Zhang, J. & Rivest, S. *Eur. J. Neurosci.* **11**, 2651–2668 (1999).
5. Ek, M., Arias, C., Sawchenko, P. & Ericsson-Dahlstrand, A. *J. Comp. Neurol.* **428**, 5–20 (2000).
6. Engblom, D., Ek, M., Hallbeck, M., Ericsson-Dahlstrand, A. & Blomqvist, A. *Neurosci. Lett.* **281**, 163–166 (2000).
7. Jakobsson, P. J., Thoren, S., Morgenstern, R. & Samuelsson, B. *Proc. Natl Acad. Sci. USA* **96**, 7220–7225 (1999).
8. Van Dam, A. M., Brouns, M., Man, A. H. W. & Berkenbosch, F. *Brain Res.* **613**, 331–336 (1993).
9. Thoren, S. & Jakobsson, P. J. *Eur. J. Biochem.* **267**, 6428–6434 (2000).
10. Murakami, M. *et al.* *J. Biol. Chem.* **275**, 32783–32792 (2000).
11. Ericsson, A., Kovacs, K. J. & Sawchenko, P. E. *J. Neurosci.* **14**, 897–913 (1994).

Pollination

Flexible style that encourages outcrossing

Despite the convenience of self-pollination (selfing) in flowering plants^{1–3}, the detrimental effects of inbreeding that follow repeated selfing^{3,4} have promoted strong natural selection for mating systems that ensure successful cross-fertilization (outcrossing). Here we describe a mechanism deployed by some tropical ginger flowers to avoid self-pollination — the flower moves its stigma (style), which normally acts as the pollen receptor, out of the way while its anther is releasing pollen. This cunning evasion adds to the diversity of pollination strategies that have contributed to the evolutionary success of flowering plants.

Alpinia is an Asian genus in the ginger family (Zingiberaceae) containing more than 250 species⁵. These are perennials with terminal inflorescences that produce between two and ten open flowers every day; each flower is hermaphrodite and lasts for only a day. We have monitored how the flower parts behave in nine species of *Alpinia*, both native and introduced, in a tropical seasonal rainforest in Xishuangbanna, Yunnan, in southwest China⁶.

Each species of *Alpinia* has two phenotypes that coexist in all populations and which differ in the movement of the flower stigma (the phenotypes are termed cataflexistyled or hyperflexistyled flowers, depending on the direction of stigma movement during flowering). When cataflexistyled flowers are fully open (06:00–06:30), the stigma is held above the open (dehiscence) anther from which pollen is being released (Fig. 1a). At the same time of day, the receptive stigma of hyperflexistyled flowers is curved downwards, below the indehiscent anther from which pollen has not yet been shed (Fig. 1b).

Flowers of both types retain these respective stigma positions until about mid-day, when the stigma of the hyperflexistyled form elongates and becomes erect above the anther (male phase). This movement prevents contact with insect visitors and creates an angle larger than 170° between the stigma and the anther's ventral face (11:45–13:30); the anther then dehiscence and pollen is released (14:30–15:00; Fig. 1d).

The movement of the style of the cataflexistyled form is slower: here the stigma begins to move downwards and enter the receptive position (female phase; less than 170° from the anther's dorsal face) between 14:40 and 15:00 (several minutes after anther dehiscence in hyperflexistyled flowers; Fig. 1c). Flowering (or anthesis) ends in both forms after dark, when the anthers collapse and the corolla flops down.

The speed of stylar movement depends

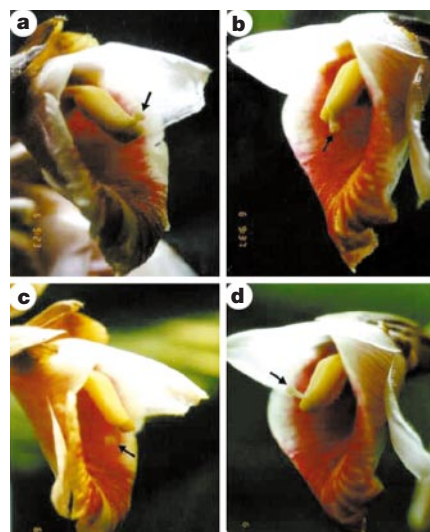


Figure 1 Positions of the stigma of the two flower forms in *Alpinia kwangsiensis* at different stages of flowering. **a**, Cataflexistyled flower in its male phase (before noon), in which the stigma is reflexed above the dehiscence anther. **b**, Hyperflexistyled flower in its female phase (before noon), in which the stigma is deflexed below the indehiscent anther. **c**, The same flower as in **a** during its female phase (afternoon), with the stigma below the anther; note that pollen has been removed from the anther by insect visitors (mainly xylocopid bees). **d**, The same flower as in **b**, but in its male phase (afternoon), with the stigma now erect above the anther, which then sheds its pollen. Arrows, stigma position.

on the weather conditions, but all flowers of the same phenotype that open on the same day are strictly synchronous. The anthers of the hyperflexistyled flower never dehiscence before all stigmas of the same phenotype have moved out of the receptive position (Fig. 2). It is likely that successful pollination only occurs between the two different forms, with the two phenotypes being associated with two genotypes (for example, in a natural population of *A. kwangsiensis* the ratio of individuals of the two phenotypes is about unity: 86:78; $\chi^2 = 0.39$, $P > 0.5$).

We artificially manipulated different pollination combinations within and between phenotypes of *A. kwangsiensis* in the field. Our results indicate that fruit set resulting from cross-pollination between the two phenotypes is not significantly different ($F = 1.393$, d.f. = 1, $P = 0.242$) and that for the same treatments (self-pollination, cross-pollination, open pollination or controls), fruit-set rates did not differ significantly between the two phenotypes ($F = 2.251$, d.f. = 4, $P = 0.072$), indicating self-compatibility of the species. However, there was a significant difference between the treatments within the same phenotype ($F = 69.163$, d.f. = 6, $P < 0.001$): in both forms and during both gender phases, cross-pollination had a significantly higher fruit set than self-pollination, indicative of an inbreeding depression effect.

The floral strategy described here not only prevents self-pollination in a flower and within the same individual, but also

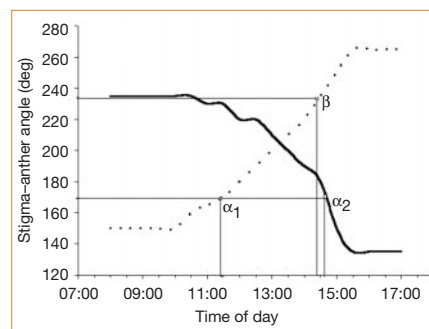


Figure 2 Floral behaviour of *Alpinia kwangsiensis* during a single day of flowering. There is no overlap in the male and female phases of the two phenotypes (dotted line, cataflexistyled flower; full line, hyperflexistyled flower). α_1 , Time when the stigma of a hyperflexistyled flower becomes reflexed out of its receptive position; α_2 , time when the stigma of a cataflexistyled flower begins to move downwards; β , time when the anther dehiscence of the hyperflexistyled flower.

among individuals of the same phenotype. It decreases inbreeding and promotes outcrossing in the plant by temporally and spatially separating the presentation of pollen and receptive stigmas through active floral movement. This mechanism, which we call flexistylis, differs from other passive outbreeding devices, such as dichogamy, herkogamy, enantiostylis and heterostylis⁷, in that it combines some features of all of these mechanisms with the unique movement of floral parts.

We observed flexistylis in all nine *Alpinia* species we studied⁸. In a molecular analysis of the phylogenetic relationships within the Zingiberaceae family (W. J. K. *et al.*, unpublished data), these nine species are distributed in three separate clades in the Alpineae, indicating that flexistylis either evolved independently several times in this Alpineae group or that it is widespread (though as yet unrecorded) in many taxa in the group (in *Amomum*, for example⁹).

Qing-Jun Li*, **Zai-Fu Xu***, **W. John Kress†**, **Yong-Mei Xia***, **Ling Zhang***, **Xiao-Bao Deng***, **Jiang-Yun Gao***, **Zhi-Lin Bai***

*Xishuangbanna Tropical Botanical Garden, The Chinese Academy of Sciences, Mengla, Yunnan 666303, China

e-mail: qjlixtbg@bn.yn.cninfo.net

†Department of Botany, National Museum of Natural History, Smithsonian Institution, Washington DC 20560-0166, USA

‡Kunming Institute of Botany, The Chinese Academy of Sciences, Kunming, Yunnan 650204, China

1. Baker, H. G. *Evolution* **9**, 347–348 (1955).
2. Stebbins, G. L. *Am. Nat.* **91**, 337–354 (1957).
3. Darwin, C. *The Effects of Cross- and Self-fertilization in the Vegetable Kingdom* 2nd edn (Murray, London, 1916).
4. Holsinger, K. E. *Trends Ecol. Evol.* **6**, 307–308 (1991).
5. Smith, R. M. *Edinbr. J. Bot.* **47**, 1–75 (1990).
6. Zhang, J.-H. & Cao, M. *Biol. Conserv.* **73**, 229–238 (1995).
7. Richards, A. J. *Plant Breeding Systems* 2nd edn (Chapman & Hall, London, 1997).
8. Li, Q.-J., Xu, Z.-F. & Xia, Y.-M. *Acta Botanica Sinica* (in the press).
9. Cui, X.-L., Wei, R.-C. & Huang, R.-F. in *Proc. 2nd Symp. Fam. Zingiberaceae* (eds Wu, T.-L., Wu, Q.-G. & Chen, Z.-Y.) 288–296 (Zhongshan Univ. Press, Guangzhou, China, 1996).

New hominin genus from eastern Africa shows diverse middle Pliocene lineages

Meave G. Leakey*, Fred Spoor†, Frank H. Brown‡, Patrick N. Gathogo‡, Christopher Kiarie*, Louise N. Leakey* & Ian McDougall§

* Division of Palaeontology, National Museums of Kenya, P.O. Box 40658, Nairobi, Kenya

† Department of Anatomy & Developmental Biology, University College London, WC1E 6JJ, UK

‡ Department of Geology & Geophysics, University of Utah, Salt Lake City, Utah 84112, USA

§ Research School of Earth Sciences, The Australian National University, Canberra ACT 0200, Australia

Most interpretations of early hominin phylogeny recognize a single early to middle Pliocene ancestral lineage, best represented by *Australopithecus afarensis*, which gave rise to a radiation of taxa in the late Pliocene. Here we report on new fossils discovered west of Lake Turkana, Kenya, which differ markedly from those of contemporary *A. afarensis*, indicating that hominin taxonomic diversity extended back, well into the middle Pliocene. A 3.5 Myr-old cranium, showing a unique combination of derived facial and primitive neurocranial features, is assigned to a new genus of hominin. These findings point to an early diet-driven adaptive radiation, provide new insight on the association of hominin craniodental features, and have implications for our understanding of Plio–Pleistocene hominin phylogeny.

The eastern African hominin record between 4 and 3 Myr is represented exclusively by a single species, *A. afarensis*, and its possible ancestor, *Australopithecus anamensis*, which are commonly thought to belong to the lineage ancestral to all later hominins^{1,2}. This apparent lack of diversity in the middle Pliocene contrasts markedly with the increasingly bushy phylogeny evident in the later hominin fossil record. To study further the time interval between 4 and 3 Myr, fieldwork in 1998 and 1999 focused on sites of this age at Lomekwi in the Nachukui Formation, west of Lake Turkana. New hominin discoveries from Lomekwi, as well as two mandibles and isolated molars recovered previously³ (Table 1), indicate that multiple species existed between 3.5 and 3.0 Myr. The new finds include a well-preserved temporal bone, two partial maxillae, isolated teeth, and most importantly a largely complete, although distorted, cranium. We assign the latter specimen to a new hominin genus on the basis of its unique combination of primitive and derived features.

Description of *Kenyanthropus platyops*

Order Primates LINNAEUS 1758

Suborder Anthropoidea MIVART 1864

Superfamily Hominoidea GRAY 1825

Kenyanthropus gen. nov.

Etymology. In recognition of Kenya's contribution to the understanding of human evolution through the many specimens recovered from its fossil sites.

Generic diagnosis. A hominin genus characterized by the following morphology: transverse facial contour flat at a level just below the nasal bones; tall malar region; zygomaticoalveolar crest low and curved; anterior surface of the maxillary zygomatic process positioned over premolars and more vertically orientated than the nasal aperture and nasoalveolar clivus; nasoalveolar clivus long and both transversely and sagittally flat, without marked jugal; moderate subnasal prognathism; incisor alveoli parallel with, and only just anterior to, the bicanine line; nasal cavity entrance stepped; palate roof thin and flexed inferiorly anterior to the incisive foramen; upper incisor (I^1 and I^2) roots near equal in size; upper premolars (P^3 , P^4) mostly three-rooted; upper first and second molars (M^1 and M^2) small with thick enamel; tympanic element mediolaterally long and lacking a petrous crest; external acoustic porus small. *Kenyanthropus* can be distinguished from *Ardipithecus ramidus* by its buccolingually narrow M^2 , thick molar enamel, and a temporal

bone with a more cylindrical articular eminence and deeper mandibular fossa. It differs from *A. anamensis*, *A. afarensis*, *A. africanus* and *A. garhi* in the derived morphology of the lower face, particularly the moderate subnasal prognathism, sagittally and transversely flat nasoalveolar clivus, anteriorly positioned maxillary zygomatic process, similarly sized I^1 and I^2 roots, and small M^1 and M^2 crowns. From *A. afarensis* it also differs by a transversely flat midface, a small, external acoustic porus, and the absence of an occipital/marginal venous sinus system, and from *A. africanus* by a tall malar region, a low and curved zygomaticoalveolar crest, a narrow nasal aperture, the absence of anterior facial pillars, a tubular, long and crestless tympanic element, and a small, external acoustic porus. *Kenyanthropus* lacks the suite of derived dental and cranial features found in *Paranthropus aethiopicus*, *P. boisei* and *P. robustus* (Table 2), and the derived cranial features of species indisputably assigned to *Homo* (For example, *H. erectus* s.l. and *H. sapiens*, but not *H. rudolfensis* and *H. habilis*)⁴.

Type species *Kenyanthropus platyops* sp. nov.

Etymology. From the Greek *platys*, meaning flat, and *opsis*, meaning face; thus referring to the characteristically flat face of this species.

Specific diagnosis. Same as for genus.

Types. The holotype is KNM-WT 40000 (Fig. 1a–d), a largely complete cranium found by J. Erus in August 1999. The paratype is KNM-WT 38350 (Fig. 1e), a partial left maxilla found by B. Onyango in August 1998. The repository is the National Museums of Kenya, Nairobi.

Localities. Lomekwi localities are situated in the Lomekwi and Topernawi river drainages in Turkana district, northern Kenya (Fig. 2). The type locality LO-6N is at 03° 54.03' north latitude, 035° 44.40' east longitude.

Horizon. The type specimen is from the Kataboi Member, 8 m below the Tulu Bor Tuff and 12 m above the Lokochot Tuff, giving an estimated age of 3.5 Myr. The paratype is from the lower Lomekwi Member, 17 m above the Tulu Bor Tuff, with an estimated age of 3.3 Myr.

Cranial description and comparisons

The overall size of the KNM-WT 40000 cranium falls within the range of *A. afarensis* and *A. africanus*. It is preserved in two main parts, the neurocranium with the superior and lateral orbital margins, but lacking most of the cranial base; and the face, lacking

the premolar and anterior tooth crowns and the right incisor roots. Most of the vault is heavily distorted, both through post-mortem diploic expansion and compression from an inferoposterior direction (Fig. 1a, b). The better preserved facial part shows some lateral skewing of the nasal area, anterior displacement of the right canine,

and some expansion of the alveolar and zygomatic processes (Fig. 1c–d), but allows for reliable assessment of its morphology.

Only the right M² crown is sufficiently preserved to allow reliable metric dental comparisons. It is particularly small, falling below the known ranges of other early hominin species (Fig. 3a). Likewise, the

Table 1 Hominin specimens from the lower Lomekwi and Kataboi Members

KNM-WT	Description	Year	Discoverer	Locality	Measurements (mm)
8556	Mandible fragment: symphysis, right body with RP ₃ –RM ₁ , isolated partial RM ₂ , RM ₃ , LP ₃	1982	N. Mutiwa	LO-5	RP ₃ , 9.8, 12.4; RP ₄ , 11.3, 12.6; RM ₁ , 13.7, 12.9; RM ₂ , NA, NA; RM ₃ , (17.5), (14.1); LP ₃ , 9.8, 12.5
8557	LM _{1/2}	1982	N. Mutiwa	LO-4	NA, (11.5)
16003	RM ³	1985	M. Kyeva	LO-5	13.3, 14.6
16006	Left mandible fragment with M ₂ fragment and M ₃	1985	N. Mutiwa	LO-4E	M ₂ , NA, NA; M ₃ , 15.3, 13.1
38332	Partial RM ³ crown	1999	M. Eregae	LO-4E	NA, 14.8
38333	LM _{1/2} crown	1999	M. Eregae	LO-4E	13.1, 12.1
38334	LM _{1/2}	1999	M. Eregae	LO-4W	12.1, 11.5
38335	RM _{1/2} crown fragment	1999	M. Eregae	LO-4E	NA
38337	RM _{1/2}	1999	R. Moru	LO-4E	11.5, 12.3
38338	Partial RM ^{1/2} crown	1999	N. Mutiwa	LO-4E	NA
38339	LM _{1/2} crown	1999	J. Erus	LO-4W	12.8, 12.7
38341	Partial LM _{2/3}	1999	G. Ekalale	LO-4E	NA
38342	LM _{1/2} crown	1999	J. Erus	LO-4E	12.8, (11.3)
38343	Right maxilla fragment with I ² and P ³ roots and partial C; mandible fragment with partial P ₄ and M ₁ roots	1999	J. Erus	LO-4W	NA
38344	RM _{1/2} crown	1998	M. Eregae	LO-9	12.8, 12.2
38346	Partial RM ^{1/2}	1998	M. Mutiwa	LO-5	NA
38347	LdM ₂ crown	1998	R. Moru	LO-5	11.7, 9.6
38349	RM _{1/2} crown	1998	W. Mangao	LO-5	13.5, 12.6
38350	Left maxilla fragment with P ³ and P ⁴ roots and partial M ¹	1998	B. Onyango	LO-5	LM ¹ : (10.5), (12.0)
38352	Partial RM _{1/2}	1998	W. Mangao	LO-5	NA, 11.5
38355	Partial RM ^{1/2} crown	1998	M. Eregae	LO-9	NA
38356	Partial RM ^{1/2} crown	1998	M. Eregae & J. Kaatho	LO-9	12.8, NA
38357	RM _{1/2}	1998	G. Ekalale	LO-5	12.8, 11.8
38358	Associated RI ² , LM ₂ fragment, LM ₃ , RM ³ fragment, four crown fragments	1998	G. Ekalale	LO-5	RI ² , 7.5, 7.5, 9.1; LM ₃ , 15.3, 13.2
38359	Associated RM ₁ , RM ₂	1998	M. Eregae	LO-5	RM ₁ , 12.7, 11.6; RM ₂ , 13.9, 12.2
38361	Associated (partial) germs of I ¹ , LI ² , RC, LRP ³ , LRP ⁴	1998	R. Moru	LO-5	I ¹ , NA, (8.0), (11.5); LI ² , 7.6, >5.9, 8.3; LP ³ , (9.3), (12.0)
38362	Associated partial LM ^{1/2} , RM ^{1/2}	1998	R. Moru	LO-5	RM ^{1/2} , 12.9, 14.3
39949	Partial LP ₄	1998	R. Moru	LO-5	NA
39950	RM ₃	1998	R. Moru	LO-5	16.0, 14.5
39951	RM _{1/2} fragment	1998	R. Moru	LO-5	NA
39952	LM _{1/2}	1998	R. Moru	LO-5	NA
39953	LM _{1/2} fragment	1998	R. Moru	LO-5	NA
39954	Two tooth fragments	1998	R. Moru	LO-5	NA
39955	L _C fragment	1998	R. Moru	LO-5	NA
40000	Cranium	1999	J. Erus	LO-6N	RM ² , 11.4, 12.4
40001	Right temporal bone	1998	P. Gathogo	LO-5	NA

Dental measurements taken as in ref. 34. Mesiodistal crown diameter followed by buccolingual or labiolingual diameter, and for incisors and canines, labial crown height. Values in parentheses are estimates. NA, Not available. L or R in the 'Description' column indicates the left or right side. C, upper canine; d, deciduous.

Table 2 Derived cranial features of *Paranthropus*, and their character state in *K. platyops* and *H. rudolfensis*

	<i>Paranthropus aethiopicus</i>	<i>Paranthropus boisei</i>	<i>Paranthropus robustus</i>	<i>Kenyanthropus platyops</i>	<i>Homo rudolfensis</i>
Upper molar size	Large	Large	Moderate	Small	Moderate
Enamel thickness	Hyperthick	Hyperthick	Hyperthick	Thick	Thick
Palatal thickness	Thick	Thick	Thick	Thin	Thin
Incisor alveoli close to bicanine line*	Present	Present	Present	Present	Present
Nasoalveolar clivus	Gutter	Gutter	Gutter	Flat	Flat
Midline subnasal prognathism	Strong	Moderate	Moderate	Weak	Weak
Upper I ² root to lateral nasal aperture	Medial	Medial	Medial	Lateral	Lateral
Nasal cavity entrance	Smooth	Smooth	Smooth	Stepped	Stepped
Zygomaticoalveolar crest	Straight, high	Straight, high	Straight, high	Curved, low	Curved, low
Anteriorly positioned zygomatic process of maxilla*	Present	Present	Present	Present	Present
Midface transverse contour	Concave, dished	Concave, dished	Concave, dished	Flat	Flat
Malar region	Wide	Wide	Wide	Tall	Tall
Malar orientation to lateral nasal margin	Aligned	Aligned	Aligned	More vertical	More vertical
Facial hafting, frontal trigone	High, present	High, present	High, present	Low, absent	Low, absent
Postorbital constriction	Marked	Marked	Marked	Moderate	Moderate
Initial supraorbital course of temporal lines	Medial	Medial	Medial	Posteromedial	Posteromedial
Tympanic vertically deep and plate-like	Present	Present	Present	Absent	Absent
Position external acoustic porus	Lateral	Lateral	Lateral	Medial	Medial
Mandibular fossa depth	Shallow	Deep	Deep	Moderate	Moderate
Foramen magnum heart shaped	Present	Present	Absent	Absent	Absent
Occipitomarginal sinus	Unknown	Present	present	Absent	Absent

Hypodigm of *H. rudolfensis* as in ref. 35. See refs 1, 8, 11, 36–40 for detailed discussions of the features.

* Character states shared by *Paranthropus* and *K. platyops*.

estimated M^1 crown size of KNM-WT 38350 (Table 1) corresponds to minima for *A. anamensis*, *A. afarensis* and *H. habilis*, and is below the ranges for other African early hominins^{3–7}. Molar enamel thickness in both specimens is comparable to that in *A. anamensis* and *A. afarensis*. CT scans show that both P^3 and P^4 of KNM-WT 40000 have a lingual root and two well-separated buccal roots. This morphology, thought to be the ancestral hominoid condition⁸, is commonly found in *Paranthropus*, but is variable among species of *Australopithecus*. The P^3 of KNM-WT 38350 has three well-separated roots (Fig. 1e). Its P^4 seems to be two-rooted, but the deeply grooved buccal root may split more apically. Relative to M^2 crown size, the canine roots of KNM-WT 40000 are smaller in cross-section at the alveolar margin than in *Ardipithecus ramidus* and *A. anamensis*, similar in size to *A. afarensis*, *A. africanus* and *H. habilis*, and larger than in *P. boisei*. Exposed surfaces and CT scans demonstrate that the I^1 and I^2 roots in KNM-WT 40000 are straight and similar in size. At the level of the alveolar margin the cross-sectional area of the I^2 root is about 90% of that of the I^1 root, whereas this is typically 50–70% in other known hominid taxa.

The incisor alveoli of KNM-WT 40000 are aligned coronally, just anterior to the bicanine line, and the overlying nasoalveolar clivus is flat both sagittally and transversely. There is no canine jugum visible on the preserved left side, reflecting the modest size of the canine root. At 32 mm (chord distance nasospinale to prosthion) the clivus is among the longest of all early hominins. Subnasal prognathism is

moderate, expressed by a more vertically orientated clivus than in nearly all specimens of *Australopithecus* and *Paranthropus* (Fig. 3b). The nasal aperture lies in the same coronal plane as the nasoalveolar clivus and there are no anterior facial pillars (Fig. 1a–c). The nasal aperture is small and narrow, in contrast to the large, wide aperture in *A. africanus* and *P. robustus*. The midface of KNM-WT 40000 is dominated by the tall malar region (Fig. 3c) with a low and curved zygomaticoalveolar crest. At a level just below the nasal bones the transverse facial contour is flat (Fig. 1b). In both KNM-WT 40000 and KNM-WT 38350 the anterior surface of the zygomatic process of the maxilla is positioned between P^3 and P^4 (Fig. 1a, d, e), as is commonly seen in *Paranthropus*, but more anteriorly than in most *Australopithecus* specimens⁹ or in *H. habilis*. The supraorbital region is *Australopithecus*-like, lacking both a frontal trigon as seen in *Paranthropus*, and a supratatorial sulcus as seen in *H. habilis* (but not *H. rudolfensis*). Relative postorbital constriction (frontofacial index) of KNM-WT 40000 is similar to that in *Australopithecus*, *H. rudolfensis* and *H. habilis*, and less than in *P. boisei* (estimated frontofacial index⁹ = 70). Its temporal lines converging on the frontal squama have a posteromedial course throughout (Fig. 1b). Around bregma the midline morphology is not well preserved, but the posterior half of the parietals show double, slightly raised temporal lines about 6 mm apart. These contribute posteriorly to indistinct compound temporal/nuchal lines. The original shape of the severely distorted mastoids cannot be reconstructed, but other

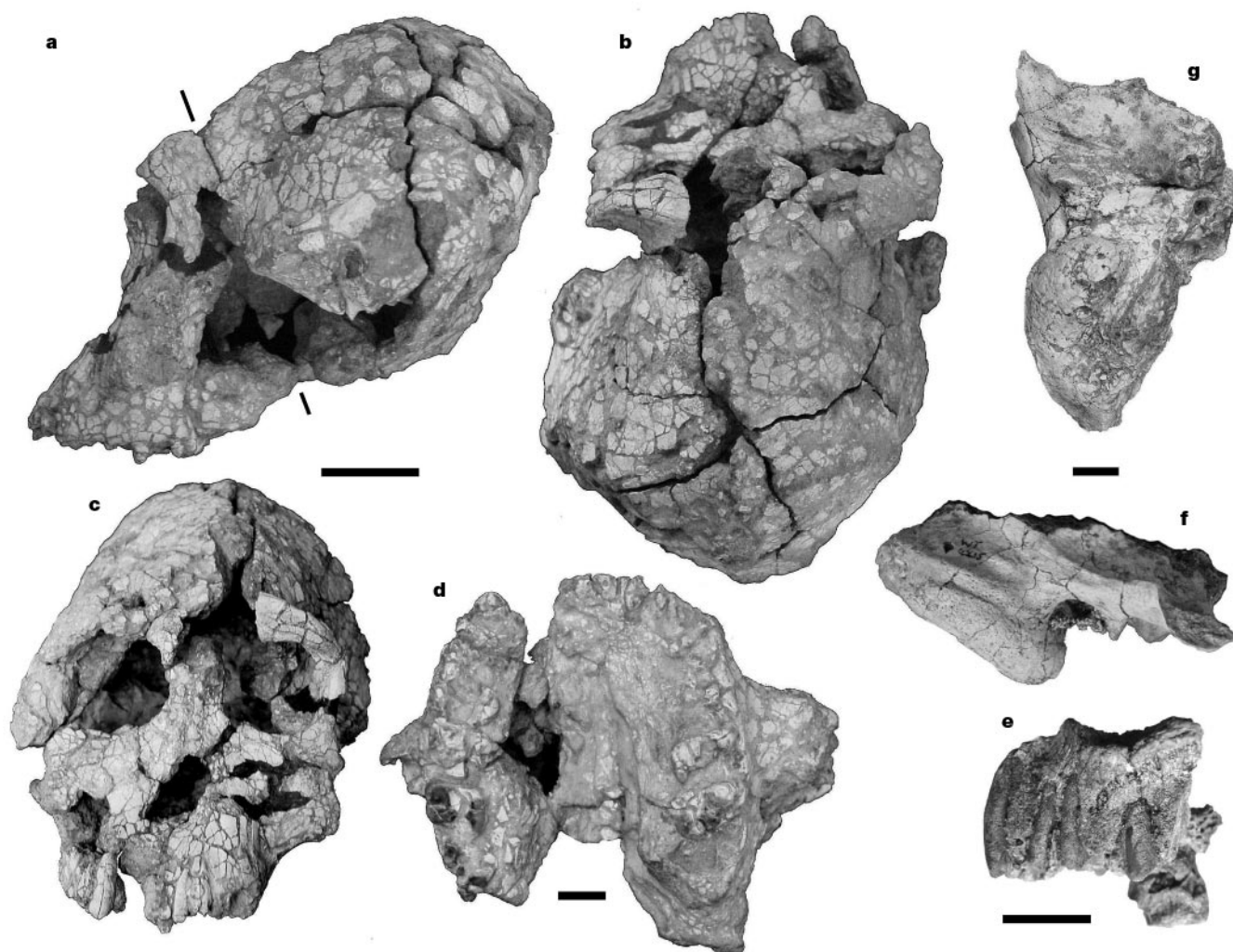


Figure 1 Holotype KNM-WT 40000 **a**, left lateral view (markers indicate the plane separating the distorted neurocranium and the well-preserved face). **b**, Superior view.

c, Anterior view. **d**, Occlusal view of palate. Paratype KNM-WT 38350. **e**, Lateral view. KNM-WT 40001. **f**, Lateral view. **g**, Inferior view. Scale bars: **a–c**, 3 cm; **d–g**, 1 cm.

parts of the left temporal are well preserved. The tubular tympanic lacks a petrous crest and forms a narrow external acoustic meatus with a small aperture. This combination constitutes the primitive hominin morphology, also seen in *Ar. ramidus* and *A. anamensis* (Fig. 3d). The mandibular fossa resembles that of specimens of *A. afarensis* and *A. africanus*. It is moderately deep, and the articular eminence, missing its lateral margin, is cylindrical with a moderately convex sagittal profile. The preserved posterior half of the foramen magnum suggests that it was probably oval in shape, rather than the heart shape seen in *P. boisei* and probably *P. aethiopicus*. Regarding the endocranial aspect, the reasonably well preserved occipital surface lacks any indication of the occipital/marginal venous sinus system characteristic of *A. afarensis*, *P. boisei* and *P. robustus*. Bilateral sulci suggest that the transverse/sigmoid sinus system was well developed. Endocranial capacity is difficult to estimate because of the distorted vault. However, comparing hominin glabella–opisthion arc lengths⁸ with that of KNM-WT 40000 (259 mm; an estimate inflated by diploic expansion) suggests a value in the range of *Australopithecus* or *Paranthropus*.

The sex of KNM-WT 40000 is difficult to infer. Interpretation of the canine root size proves inconclusive without a suitable comparative context. The small M² crown size could suggest that the specimen is female. However, the close proximity and slightly raised aspect of the temporal lines on the posterior half of the parietals is not seen in known female hominin crania, including the *Paranthropus* specimens KNM-ER 732, KNM-ER 407 and DNH7, and suggests that KNM-WT 40000 could be male.

With incisor alveoli close to the bicanine line and anteriorly positioned zygomatic processes, the face of KNM-WT 40000 resembles the flat, orthognathic-looking faces of both *Paranthropus* and *H. rudolfensis* cranium KNM-ER 1470. However, KNM-WT 40000 lacks most of the derived features that characterize *Paranthropus* (Table 2), and its facial architecture differs from the latter in much the same way as has been described for KNM-ER 1470 (refs 8, 10). Facial flatness in *Paranthropus* results from the

forward position of the anteroinferiorly sloping malar region, whose main facial surface approximates the plane of the nasal aperture, but whose orientation contrasts with the more horizontally inclined nasoalveolar gutter¹¹. In KNM-WT 40000 and KNM-ER 1470, it is the flat and orthognathic nasoalveolar clivus that aligns with the plane of the nasal aperture, whereas the anteriorly set, tall malar region is more vertically orientated. KNM-WT 40000 lacks the derived short nasal bones and everted lateral nasal margin of KNM-ER 1470, and is less orthognathic in the midfacial region than this specimen; however, on balance this is the hominin face that KNM-WT 40000 most closely resembles.

Additional material

The right maxilla fragment KNM-WT 38343A preserves three well-separated P³ roots, and its damaged canine seems low-crowned when compared with *A. afarensis* canines of similar size and degree of wear. The right temporal bone KNM-WT 40001 lacks the squama and petrous apex, but is otherwise well preserved (Fig. 1f, g). It shows a combination of characters not seen in any other hominin specimen. The projecting mastoid process is rounded, with an anteriorly positioned tip. It has a well-developed digastric fossa in the form of a deep, narrow groove that runs posterolaterally from the stylomastoid foramen, fully demarcating the mastoid process from the adjacent nuchal plane. The tympanic element is long, inferosuperiorly shallow and lacks a petrous crest. The external acoustic porus is the smallest of any known hominin temporal bone (Fig. 3d). The articular eminence is as broad mediolaterally (38 mm) as in *P. aethiopicus* and *P. boisei*, and similar to the largest found in *A. afarensis*. Compared with KNM-WT 40000 the eminence is relatively flat sagittally, and the mandibular fossa is shallow.

The partial mandibles KNM-WT 8556 and KNM-WT 16006 have been assigned to *A. afarensis*³. However, KNM-WT 8556 shows a more derived morphology than this species by having a flat, more horizontal post-incisive plane, a more superiorly positioned genio-glossal pit, a molarized lower fourth premolar (P₄) and a large M₃

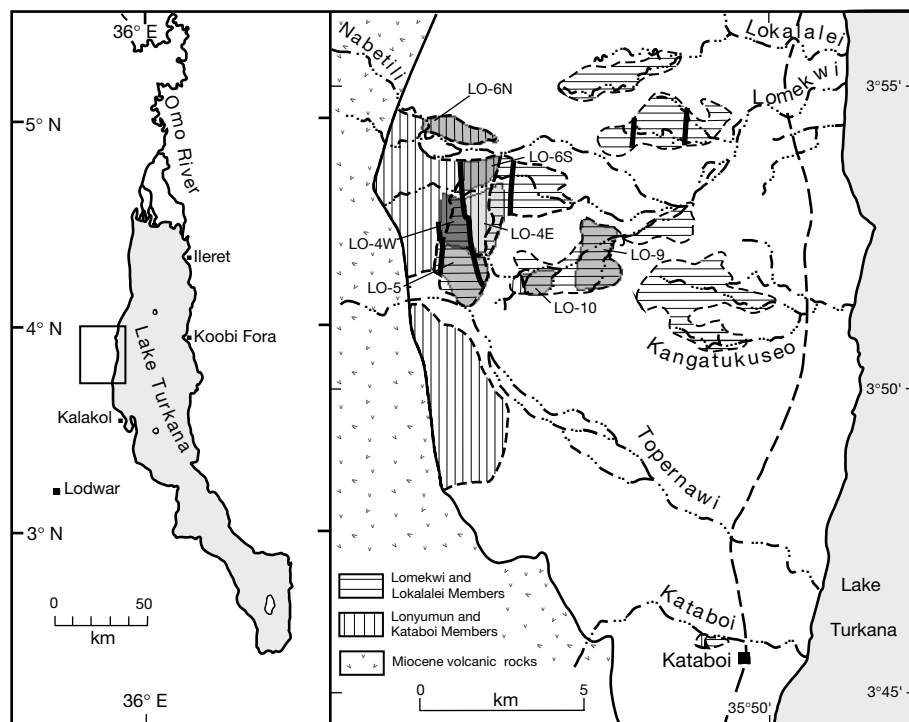


Figure 2 Map showing localities of fossil collection in upper Lomekwi and simplified geology. The boundary between the Kataboi and Lomekwi Members is the base of the Tulu Bor Tuff, indicated as a dashed line through LO-4E and LO-4W. Faults are shown as

thick lines; minor faults are omitted. LO-4E and LO-4W are of different shades to distinguish them from each other.

(ref. 3). Indeed, relative to its *Australopithecus*-sized M_1 (refs 5, 6, 12, 13), the P_4 and M_3 crowns of KNM-WT 8556 are enlarged to an extent only seen in *P. boisei* (Fig. 3e, f). All unworn molars in the Lomekwi sample are characterized by low occlusal relief and numerous secondary fissures. Most of the lower molars, including the KNM-WT 16006 M_3 , have a well-developed protostylid, a feature that is usually absent in *A. afarensis*, but common in *A. africanus*¹⁴. The two I_2 s are lower crowned than in *A. afarensis*, *A. africanus*¹⁴ and *P. robustus*¹⁴. Inability to distinguish between first and second molars makes meaningful intertaxon comparisons of these elements difficult.

Taxonomic discussion

The hominin specimens recovered from the Kataboi and lower Lomekwi Members show a suite of features that distinguishes them from established hominin taxa, including the only contemporaneous eastern African species, *A. afarensis*. Compared with the latter, the morphology of *K. platyops* is more derived facially, and more primitive in its small external acoustic porus and the absence of an occipital/marginal sinus system. These finds not only provide evidence for a taxonomically more diverse middle Pliocene hominin record, but also show that a more orthognathic facial morphology emerged significantly earlier in hominin evolutionary history than previously documented. This early faciodental diversity concerns morphologies that functionally are most closely associated with mastication. It suggests a diet-driven adaptive radiation among hominins in this time interval, which perhaps had its origins

considerably earlier. Furthermore, the presence in *K. platyops* of an anteriorly positioned zygomatic process in combination with a small M^1 and M^2 indicates that such characters are more independent than is suggested by developmental and functional models that link such facial morphology in *Paranthropus* with postcanine megadontia^{11,15}.

At present it is unclear whether the Lomekwi hominin fossils sample multiple species. Apart from the paratype maxilla KNM-WT 38350 with its small molar size and anteriorly positioned zygomatic process, the other specimens cannot be positively associated morphologically with the *K. platyops* holotype. These are therefore not included in the paratype series, and are left unassigned until further evidence emerges. Differences between the tympanic and mandibular fossa morphologies of the KNM-WT 40000 and KNM-WT 40001 temporal bones can perhaps be accommodated within a single species, but their shared primitive characters do not necessarily imply conspecificity. Affiliation of the KNM-WT 8556 mandible with the *K. platyops* types is not contradicted by its molarized P_4 , which is consistent with an anteriorly positioned zygomatic process. However, its M_1 is larger than would be inferred from the smaller upper molars of the types, and with a 177 mm² crown area it is also larger than any in the combined sample of ten isolated M_1 s and M_2 s (139–172 mm²). One isolated $M^{1/2}$ (KNM-WT 38362) is significantly larger than the molars of the *K. platyops* types, whereas another (KNM-WT 38337) is similar in size to the holotype's M^2 .

The marked differences of the KNM-WT 40000 cranium from established hominin taxa, both with respect to individual features

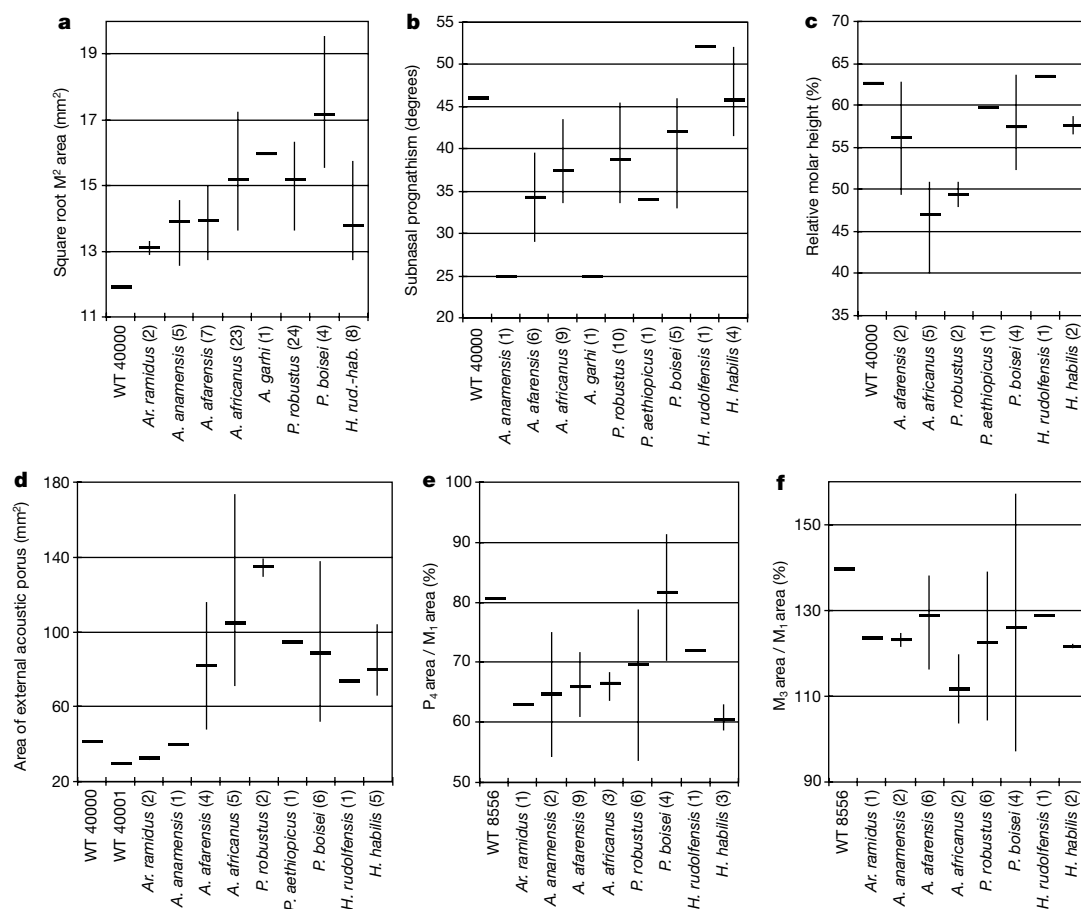


Figure 3 Mean and range of characters of specified hominins. **a**, Square root of M^2 crown area (buccolingual $\times$ mesiolingual diameters). **b**, Angle of subnasal prognathism (nasospinale–prosthion to postcanine alveolar plane). **c**, Malar height⁸ relative to orbitoalveolar height (orbitale to alveolar margin aligned with malar surface). **d**, Area of the

external acoustic porus ($\pi \times$ long axis $\times$ short axis). **e**, Crown area of P_4 relative to that of $M^1 \times 100$. **f**, Crown area of M_3 relative to that of $M^1 \times 100$. All measurements are taken from originals, directly or as given in refs 8, 9, 14, 18, 37, 41–44, except for some South African crania taken from casts. Numbers in parentheses indicate sample size.

and their unique combination, fully justify its status as a separate species. It is worth noting that comparisons with *Australopithecus bahrelghazali* cannot be made directly, because this species was named on the basis of the limited evidence provided by an anterior mandible fragment¹⁶. Specific distinction of *A. bahrelghazali* from *A. afarensis* has yet to be confirmed¹⁷, and Lomekwi specimens differ from *A. bahrelghazali* in symphyseal morphology and incisor crown height.

The generic attribution of KNM-WT 40000 is a more complex issue, in the absence of consensus over the definition of the genus category⁴. The specimen lacks almost all of the derived features of *Paranthropus* (Table 2), and there are no grounds for assigning it to this genus unless it can be shown to represent a stem species. However, the fact that the facial morphology of KNM-WT 40000 is derived in a markedly different way renders this implausible. As KNM-WT 40000 does not show the derived features associated with *Homo*⁴ (excluding *H. rudolfensis* and *H. habilis*) or the strongly primitive morphology of *Ardipithecus*¹⁸, the only other available

genus is *Australopithecus*. We agree with the taxonomically conservative, grade-sensitive approach to hominin classification that for the moment accepts *Australopithecus* as a paraphyletic genus in which are clustered stem species sharing a suite of key primitive features, such as a small brain, strong subnasal prognathism, and relatively large postcanine teeth. However, with its derived face and small molar size, KNM-WT 40000 stands apart from species assigned to *Australopithecus* on this basis. All it has in common with such species is its small brain size and a few other primitive characters in the nasal, supraorbital and temporal regions. Therefore, there is no firm basis for linking KNM-WT 40000 specifically with *Australopithecus*, and the inclusion of such a derived but early form could well render this genus polyphyletic. In a classification in which *Australopithecus* also includes the 'robust' taxa and perhaps even species traditionally known as 'early *Homo*'⁴, this genus subsumes several widely divergent craniofacial morphologies. It could thus be argued that the inclusion of KNM-WT 40000 in *Australopithecus* would merely add yet another hominin species

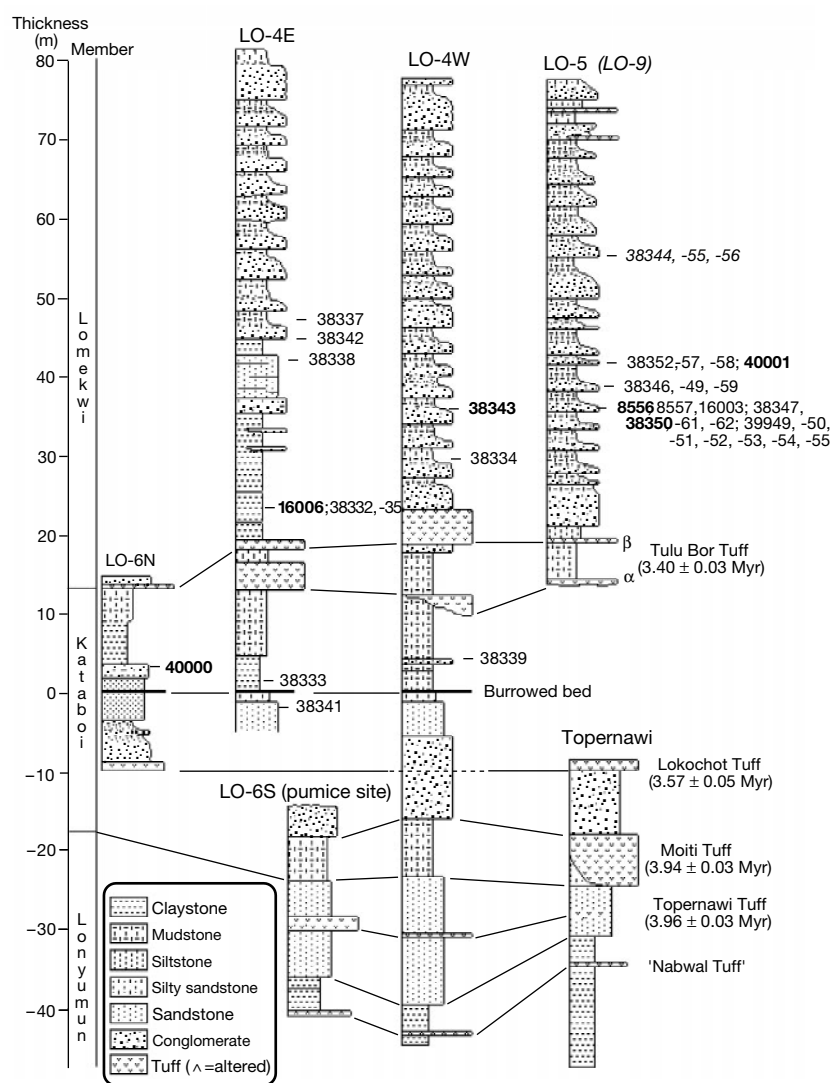


Figure 4 Stratigraphic sections and placement of hominin specimens at sites in upper part of the Lomekwi drainage, west of Lake Turkana, northern Kenya. Specimen numbers are given without the prefix KNM-WT, and those in bold are discussed in the text. Placement of specimens is relative to the nearest marker bed in the section. Italicized numbers show the relative placement of specimens at LO-9 on section LO-5. The burrowed bed, a useful local marker, is used as stratigraphic datum (0 m). Representative ⁴⁰Ar–³⁹Ar analytical data on the Moiti Tuff and on the Topernawi Tuff are given as

Supplementary Information. The date for the Tulu Bor Tuff is taken as the age of the Sidi Hakoma Tuff at Hadar²⁵, which is consistent with the age of the Toroto Tuff (3.32 ± 0.03 Myr)⁴⁵ that overlies the Tulu Bor at Koobi Fora. The age of the Lokochot Tuff is assigned from its placement at the Gilbert/Gauss Chron boundary^{24,46}. The tuff formerly thought to be the Moiti Tuff at Lomekwi²² has been informally called the 'Nabwal tuff'⁴⁷. Ages on the Tulu Bor and the Lokochot Tuffs are consistent with orbitally tuned ages of correlative ash layers in Ocean Drilling Program Core 722A in the Arabian Sea⁴⁸.

with a derived face. This amounts to defining *Australopithecus* by a single criterion, those hominin species not attributable to *Ardipithecus* or *Homo*, which in our view constitutes an undesirable approach to classification. Thus, given that KNM-WT 40000 cannot be grouped sensibly with any of the established hominin genera, and that it shows a unique pattern of facial and dental morphology that probably reflects a distinct dietary adaptive zone, we assign this specimen to the new genus *Kenyanthropus*.

Despite being separated by about 1.5 Myr, KNM-WT 40000 is very similar in its facial architecture to KNM-ER 1470, the lectotype of *H. rudolfensis*. The main differences amount to the more primitive nasal and neurocranial morphology of KNM-WT 40000. This raises the possibility that there is a close phylogenetic relationship between the two taxa, and affects our interpretation of *H. rudolfensis*. The transfer of this species to *Australopithecus* has been recommended^{4,19}, but *Kenyanthropus* may be a more appropriate genus. The identification of *K. platyops* has a number of additional implications. As a species contemporary with *A. afarensis* that is more primitive in some of its morphology, *K. platyops* weakens the case for *A. afarensis* being the sister taxon of all later hominins, and thus its proposed transfer to *Praeanthropus*^{1,20}. Furthermore, the morphology of *K. platyops* raises questions about the polarity of characters used in analyses of hominin phylogeny. An example is the species' small molar size, which, although probably a derived feature, might also imply that the larger postcanine dentition of *A. afarensis* or *A. anamensis* does not represent the primitive hominin condition. Finally, the occurrence of at least one additional hominin species in the middle Pliocene of eastern Africa means that the affiliation of fragmentary specimens can now be reassessed. For example, the attribution of the 3.3 Myr old KNM-ER 2602 cranial fragment to *A. afarensis*²¹ has been questioned⁸, and evaluating its affinities with *K. platyops* is now timely.

Geological context and dating

KNM-WT 40000 was collected near the contact of the Nachukui Formation with Miocene volcanic rocks in the northern tributary of Lomekwi (Nabetili). It is situated 12 m above the Lokochot Tuff, and 8 m below the β -Tulu Bor Tuff (Fig. 4). Along Nabetili, the Lokochot Tuff is pinkish-grey and contains much clay and volcanic detritus. It is overlain by a volcanic pebble conglomerate, followed by a pale brown quartz-rich fine sandstone that includes a burrowed fine-sandstone marker bed 10–15 cm thick. The Lokochot Tuff is replaced by a thick volcanic clast conglomerate in the central part of Lomekwi. The contact between the fine sandstone and the overlying dark mudstone can be traced from Nabetili to the hominin locality. Locally the mudstone contains volcanic pebbles at the base, and it has thin pebble conglomerate lenses in the upper part at the hominin locality, and also contains CaCO_3 concretions. The hominin specimen and other vertebrate fossils derive from this mudstone. Overlying the dark mudstone at the hominin site is a brown mudstone (8 m) that directly underlies the β -Tulu Bor Tuff.

New ^{40}Ar – ^{39}Ar determinations on alkali feldspars from pumice clasts in the Moiti Tuff and the Topernawi Tuff, stratigraphically beneath the Lokochot Tuff, were instrumental in re-investigating the lower portion of this section. The new results yield a mean age for the Topernawi Tuff of 3.96 ± 0.03 Myr; this is marginally older than the pooled age for the Moiti Tuff of 3.94 ± 0.03 Myr. Previous investigations^{22,23} placed the Topernawi Tuff above the Moiti Tuff, mainly on the basis of the K/Ar ages on alkali feldspar from pumice clasts in the Topernawi Tuff (3.78, 3.71, 3.76 and 3.97 Myr, all ± 0.04 Myr)²³. The older determination (3.97 Myr) was thought to result from contamination by detrital feldspar. South of Topernawi, however, the Topernawi Tuff has now been shown to underlie the Moiti Tuff, and to be in turn underlain by a tephra informally termed the 'Nabwal tuff', previously thought to be a Moiti Tuff correlative. The correct sequence is shown in Fig. 4, and

the new ^{40}Ar – ^{39}Ar age data on the Moiti Tuff and Topernawi Tuff are provided as Supplementary Information.

Linear interpolation between the Lokochot Tuff (3.57 Myr old)²⁴ and Tulu Bor Tuff (3.40 Myr)²⁵ yields an age of 3.5 Myr for KNM-WT 40000, and 3.53 Myr for the burrowed bed. KNM-WT 38341 from immediately below the burrowed bed has an age near 3.53 Myr. KNM-WT 38333 and 38339, from between the burrowed bed and the α -Tulu Bor Tuff lie between 3.4 and 3.5 Myr. Other specimens from LO-4, LO-5, and LO-6 lie 16–24 m above the β -Tulu Bor Tuff, with ages near 3.3 Myr. Assuming linear sedimentation between the Tulu Bor Tuff and the Lokalalei Tuff (2.5 Myr)²³, specimens from LO-9 are around 3.2 Myr. The probable error on these age estimates is less than 0.10 Myr.

Palaeogeographically, the mudstone that contained KNM-WT 40000 at LO-6N was deposited along the northern margin of a shallow lake that extended to Kataboi and beyond^{26,27}. Laterally discontinuous volcanic pebble conglomerates within the mudstone record small streams draining from hills to the west. Carbonate concretions at the hominin level are probably pedogenic, and indicate regional conditions with net evaporative loss. Other specimens between the burrowed bed and the Tulu Bor Tuff were also preserved in lake-margin environments, as is the case for KNM-WT 38341 that was collected below the burrowed bed. At LO-5, and in the upper part of LO-4E, strata were laid down by ephemeral streams draining the basin margin, principally the ancestral Topernawi, which deposited gravels in broad, shallow channels, and finer grained materials in interfluvies. Specimens preserved in floodplain deposits of the ancestral Omo River that occupied the axial portion of the basin include those at LO-9, those less than 6 m above the Tulu Bor Tuff at LO-4E, and KNM-WT 38338. Thus, there is evidence for hominins occupying floodplains of major rivers, alluvial fans, and lake-margin environments 3.0–3.5 Myr ago. There is reasonable evidence that water sources were available to these hominins in channels of the ephemeral streams, and also possibly as seeps or springs farther out into the basin.

Palaeoecology and fauna

Faunal assemblages from Lomekwi sites LO 4, LO 5, LO 6 and LO 9 indicate palaeoenvironments that were relatively well watered and well vegetated. The relative proportions of the bovids in the early collections from these sites indicate a mosaic of habitats, but with predominantly woodland and forest-edge species dominating²². Comparisons of the Lomekwi faunal assemblages with those from the few known hominin sites of similar age, Laetoli in Tanzania, Hadar in Ethiopia and Bahr el Ghazal in Chad, are of interest in view of the different hominin taxa represented. Hadar and Bahr el Ghazal, like Lomekwi, represent lakeshore or river floodplain palaeoenvironments^{28,29}, whereas Laetoli was not located near a water source; no aquatic taxa nor terrestrial mammals indicative of swamp or grassy wetlands were recovered³⁰. The faunal assemblages of all four sites indicate a mosaic of habitats that seems to have included open grasslands and more wooded or forested environments^{22,28,29,31,32}; the assemblages differ primarily in the indication of the nature of the dominant vegetation cover.

Although the mammalian faunal assemblage from Lomekwi is more similar to that from Hadar than to that from Laetoli, some mammalian species represented are different. At Lomekwi, *Theropithecus brumpti* is common and is the dominant cercopithecoid, as it is elsewhere in the Turkana Basin at this time. This species is generally considered to indicate more forested or closed woodland habitats. In the Hadar Formation, *Theropithecus darti* is the common *Theropithecus* species and is associated with lower occurrences of the water-dependent reduncines and higher occurrences of alcelaphines and/or *Aepyceros*, which indicates drier woodlands and grasslands³³. Differences in the representation of other common species at the two sites that are less obviously linked to habitat include *Kolpochoerus limnetes*, *Tragelaphus nakuae* and *Aepyceros*

shungurensis at Lomekwi, as opposed to *K. afarensis*, *T. kyalaoe* and an undescribed species of *Aepyceros* at Hadar (K. Reed, personal communication). The general indication is that the palaeoenvironment at Lomekwi may have been somewhat more vegetated and perhaps wetter than that persisting through much of the Hadar Formation. At both sites more detailed analyses will be essential to further develop an understanding of how subtle temporal changes in the faunal assemblages relate to hominin occurrences.

Note added in proof: If the hominin status of the recently published Lukeino craniodental specimens⁴⁹ is confirmed, this would support the suggestion that small molar size is the primitive rather than the derived hominin condition. □

Received 31 January; accepted 16 February 2001.

1. Strait, D. S., Grine, F. E. & Moniz, M. A. A reappraisal of early hominid phylogeny. *J. Hum. Evol.* **32**, 17–82 (1997).
2. Ward, C., Leakey, M. & Walker, A. The new hominid species *Australopithecus anamensis*. *Evol. Anthropol.* **7**, 197–205 (1999).
3. Brown, B., Brown, F. & Walker, A. New hominids from the Lake Turkana Basin, Kenya. *J. Hum. Evol.* (in the press).
4. Wood, B. A. & Collard, M. C. The human genus. *Science* **284**, 65–71 (1999).
5. Leakey, M. G., Feibel, C. S., McDougall, I. & Walker, A. C. New four-million-year-old hominid species from Kanapoi and Allia Bay, Kenya. *Nature* **376**, 565–571 (1995).
6. Leakey, M. G., Feibel, C. S., McDougall, I., Ward, C. & Walker, A. New specimens and confirmation of an early age for *Australopithecus anamensis*. *Nature* **393**, 62–66 (1998).
7. Kimbel, W. H., Johanson, D. C. & Rak, Y. Systematic assessment of a maxilla of *Homo* from Hadar, Ethiopia. *Am. J. Phys. Anthropol.* **103**, 235–262 (1997).
8. Wood, B. A. *Koobi Fora Research Project* Vol. 4 (Clarendon Press, Oxford, 1991).
9. Lockwood, C. A. & Tobias, P. V. A large male hominin cranium from Sterkfontein, South Africa, and the status of *Australopithecus africanus*. *J. Hum. Evol.* **36**, 637–685 (1999).
10. Bilsborough, A. & Wood, B. A. Cranial morphometry of early hominids: facial region. *Am. J. Phys. Anthropol.* **76**, 61–86 (1988).
11. Rak, Y. *The Australopithecine Face* (Academic, New York, 1983).
12. Lockwood, C. A., Kimbel, W. H. & Johanson, D. C. Temporal trends and metric variation in the mandibles and dentition of *Australopithecus afarensis*. *J. Hum. Evol.* **39**, 23–55 (2000).
13. Moggi-Cecchi, J., Tobias, P. V. & Beynon, A. D. The mixed dentition and associated skull fragments of a juvenile fossil hominid from Sterkfontein, South Africa. *Am. J. Phys. Anthropol.* **106**, 425–465 (1998).
14. Robinson, J. T. The dentition of the Australopithecinae. *Transv. Mus. Mem.* **9**, 1–179 (1956).
15. McCollum, M. A. The robust australopithecine face: a morphogenetic perspective. *Science* **284**, 301–305 (1999).
16. Brunet, M. *et al.* *Australopithecus bahrelghazali*, une nouvelle espèce d'Hominidé ancien de la région de Koro Toro (Tchad). *C.R. Acad. Sci. Ser. IIa* **322**, 907–913 (1996).
17. White, T. D., Suwa, G., Simpson, S. & Asfaw, B. Jaws and teeth of *A. afarensis* from Maka, Middle Awash, Ethiopia. *Am. J. Phys. Anthropol.* **111**, 45–68 (2000).
18. White, T. D., Suwa, G. & Asfaw, B. *Australopithecus ramidus*, a new species of early hominid from Aramis, Ethiopia. *Nature* **371**, 306–312 (1994).
19. Wood, B. & Collard, M. The changing face of genus *Homo*. *Evol. Anthropol.* **8**, 195–207 (2000).
20. Harrison, T. in *Species, Species Concepts, and Primate Evolution* (eds Kimbel, W. H. & Martin, L. B.) 345–371 (Plenum, New York, 1993).
21. Kimbel, W. H. Identification of a partial cranium of *Australopithecus afarensis* from the Koobi Fora Formation. *J. Hum. Evol.* **17**, 647–656 (1988).
22. Harris, J. M., Brown, F. & Leakey, M. G. Stratigraphy and paleontology of Pliocene and Pleistocene localities west of Lake Turkana, Kenya. *Cont. Sci. Nat. Hist. Mus. Los Angeles* **399**, 1–128 (1988).
23. Feibel, C. S., Brown, F. H. & McDougall, I. Stratigraphic context of fossil hominids from the Omo Group deposits, northern Turkana Basin, Kenya and Ethiopia. *Am. J. Phys. Anthropol.* **78**, 595–622 (1989).
24. McDougall, I., Brown, F. H., Cerling, T. E. & Hillhouse, J. W. A reappraisal of the geomagnetic polarity time scale to 4 Ma using data from the Turkana Basin, East Africa. *Geophys. Res. Lett.* **19**, 2349–2352 (1992).
25. Walter, R. C. & Aronson, J. L. Age and source of the Sidi Hakoma Tuff, Hadar Formation, Ethiopia. *J. Hum. Evol.* **25**, 229–240 (1993).
26. Brown, F. H. & Feibel, C. S. in *Koobi Fora Research Project*, Vol. 3, *Stratigraphy, artiodactyls and paleoenvironments* (ed. Harris, J. M.) 1–30 (Clarendon, Oxford, 1991).
27. Feibel, C. S., Harris, J. M. & Brown, F. H. in *Koobi Fora Research Project*, Vol. 3, *Stratigraphy, artiodactyls and paleoenvironments* (ed. Harris, J. M.) 321–346 (Clarendon, Oxford, 1991).
28. Johanson, D. C., Taieb, M. & Coppens, Y. Pliocene hominids from the Hadar Formation, Ethiopia

- (1973–1977): stratigraphic, chronologic and paleoenvironmental contexts, with notes on hominid morphology and systematics. *Am. J. Phys. Anthropol.* **57**, 373–402 (1982).
29. Brunet, M. *et al.* The first australopithecine 2,500 kilometres west of the Rift Valley (Chad). *Nature* **378**, 273–240 (1995).
30. Leakey, M. D. & Harris, J. M. *Laetoli, a Pliocene Site in Northern Tanzania* (Clarendon, Oxford, 1987).
31. Harris, J. M. in *Laetoli, a Pliocene Site in Northern Tanzania* (eds Leakey, M. D. & Harris, J. M.) 524–531 (Clarendon, Oxford, 1987).
32. Kimbel, W. H. *et al.* Late Pliocene *Homo* and Oldowan tools from the Hadar Formation (Kadar Hadar Member), Ethiopia. *J. Hum. Evol.* **31**, 549–561 (1996).
33. Eck, G. G. in *Theropithecus, the Rise and Fall of a Primate Genus* (ed. Jablonski, N. G.) 15–83 (Cambridge Univ. Press, 1993).
34. White, T. D. New fossil hominids from Laetoli, Tanzania. *Am. J. Phys. Anthropol.* **46**, 197–230 (1977).
35. Wood, B. Origin and evolution of the genus *Homo*. *Nature* **355**, 783–790 (1992).
36. Suwa, G. *et al.* The first skull of *Australopithecus boisei*. *Nature* **389**, 489–492 (1997).
37. Keyser, A. W. The Drimolen skull: the most complete australopithecine cranium and mandible to date. *S. Afr. J. Sci.* **96**, 189–193 (2000).
38. Kimbel, W. H., White, T. D. & Johanson, D. C. Cranial morphology of *Australopithecus afarensis*: a comparative study based on a composite reconstruction of the adult skull. *Am. J. Phys. Anthropol.* **64**, 337–388 (1984).
39. Walker, A., Leakey, R. E., Harris, J. H. & Brown, F. H. 2.5-Myr *Australopithecus boisei* from west of Lake Turkana, Kenya. *Nature* **322**, 517–522 (1986).
40. Kimbel, W. H., White, T. D. & Johanson, D. C. in *Evolutionary History of the “Robust” Australopithecines* (ed. Grine, F. E.) 259–268 (Aldine de Gruyter, New York, 1988).
41. Asfaw, B. *et al.* *Australopithecus garhi*: A new species of early hominid from Ethiopia. *Science* **284**, 629–635 (1999).
42. Grine, F. E. & Strait, D. S. New hominid fossils from Member 1 “Hanging Remnant” Swartkrans Formation, South Africa. *J. Hum. Evol.* **26**, 57–75 (1994).
43. Johanson, D. C., White, T. D. & Coppens, Y. Dental remains from the Hadar Formation, Ethiopia: 1974–1977 collections. *Am. J. Phys. Anthropol.* **57**, 545–603 (1982).
44. Tobias, P. V. in *Olduvai Gorge* Vol. 4 (Cambridge Univ. Press, 1991).
45. McDougall, I. K-Ar and ⁴⁰Ar/³⁹Ar dating of the hominid-bearing Pliocene–Pleistocene sequence at Koobi Fora, Lake Turkana, northern Kenya. *Geol. Soc. Am. Bull.* **96**, 159–175 (1985).
46. Brown, F. H., Shuey, R. T. & Croes, M. K. Magnetostratigraphy of the Shungura and Usno Formations, southwestern Ethiopia: new data and comprehensive reanalysis. *Geophys. J. R. Astron. Soc.* **54**, 519–538 (1978).
47. Haileab, B. *Geochemistry, Geochronology and Tephrostratigraphy of Tephra from the Turkana Basin, Southern Ethiopia and Northern Kenya*. Thesis, Univ. Utah (1995).
48. deMenocal, P. B. & Brown, F. H. in *Hominin Evolution and Climatic Change in Europe* Vol. 1 (eds Agustí, J., Rook, L. & Andrews, P.) 23–54 (Cambridge Univ. Press, 1999).
49. Senut, B. *et al.* First hominid from the Miocene (Lukeino Formation, Kenya). *C.R. Acad. Sci. Paris* **332**, 137–144 (2001).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank the Government of Kenya for permission to carry out this research and the National Museums of Kenya for logistical support. The National Geographic Society funded the field work and some laboratory studies. Neutron irradiations were facilitated by the Australian Institute of Nuclear Science and Engineering and the Australian Nuclear Science and Technology Organisation. We also thank the Ethiopian Ministry of Information and Culture, the National Museum of Ethiopia, B. Asfaw, Y. Bayene, C. Howell, D. Johanson, W. Kimbel, G. Suwa and T. White for permission to make comparisons with the early Ethiopian hominins and numerous people including N. Adamali, B. Asfaw, C. Dean, C. Feibel, A. Griffiths, W. Kimbel, R. Kruszynski, K. Kupczik, R. Leakey, D. Lieberman, J. Moore, K. Patel, D. Plummer, K. Reed, B. Sokhi, M. Tighe, T. White, and B. Wood for their help. Caltex (Kenya) provided fuel for the field expeditions, and R. Leakey allowed us the use of his aeroplane. The field expedition members included U. Bwana, S. Crispin, G. Ekalale, M. Eragae, J. Erus, J. Ferraro, J. Kaatho, N. Kaling, P. Kapoko, R. Lorinyok, J. Lorot, S. Hagemann, B. Malika, W. Mangao, S. Muge, P. Mulinge, D. Mutinda, K. Muthyoka, N. Mutiwa, W. Mutiwa, B. Onyango, E. Weston and J. Wynn. A. Ibui, F. Kyallo, F. Kirera, N. Malit, E. Mbua, M. Muungu, J. Ndunda, S. Ngui and A. Mwai provided curatorial assistance. The Leakey Foundation awarded a grant to E.B.

Correspondence and requests for materials should be addressed to M.G.L. (e-mail: meave@swiftkenya.com).

Observation of high-energy neutrinos using Čerenkov detectors embedded deep in Antarctic ice

E. Andrés*, P. Askebjør†, X. Bai‡, G. Barouch*, S. W. Barwick§, R. C. Bay||, K.-H. Becker¶, L. Bergström†, D. Bertrand#, D. Bierenbaum§, A. Biron*, J. Booth§, O. Botner**, A. Bouchta*, M. M. Boyce*, S. Carus††, A. Chen*, D. Chirkin||, J. Conrad**, J. Cooley*, C. G. S. Costa#, D. F. Cowen‡‡, J. Dailing§, E. Dalberg†, T. DeYoung*, P. Desiati*, J.-P. Dewulf#, P. Doksus*, J. Edsjö†, P. Ekström†, B. Erlandsson†, T. Feser§§, M. Gaug*, A. Goldschmidt|||, A. Goobar†, L. Gray*, H. Haase*, A. Hallgren**, F. Halzen*, K. Hanson‡‡, R. Hardtke*, Y. D. He, M. Hellwig§§, H. Heukenkamp*, G. C. Hill*, P. O. Hulth†, S. Hundertmark§, J. Jacobsen|||, V. Kandhadai*, A. Karle*, J. Kim§, B. Koci*, L. Köpke§§, M. Kowalski*, H. Leich*, M. Leuthold*, P. Lindahl††, I. Liubarsky*, P. Loaiza**, D. M. Lowder||, J. Ludvig|||, J. Madsen*, P. Marciniowski*, H. S. Matis|||, A. Mihalyi‡‡, T. Mikolajski*, T. C. Miller‡, Y. Minaeva†, P. Miočinić||, P. C. Mock§, R. Morse*, T. Neunhoffer§§, F. M. Newcomer‡‡, P. Niessen*, D. R. Nygren|||, H. Ögelman*, C. Pérez de los Heros**, R. Porrata§, P. B. Price||, K. Rawlins*, C. Reed§, W. Rhode†, A. Richards||, S. Richter*, J. Rodríguez Martino†, P. Romenesko*, D. Ross§, H. Rubinstein†, H.-G. Sander§§, T. Scheider§§, T. Schmidt*, D. Schneider*, E. Schneider§, R. Schwarz*, A. Silvestri*, M. Solarz||, G. M. Spiczak‡, C. Spiering*, N. Starinsky*, D. Steele*, P. Steffen*, R. G. Stokstad|||, O. Streicher*, Q. Sun†, I. Taboada‡‡, L. Thollander†, T. Thon*, S. Tilav*, N. Usechak§, M. Vander Donckt#, C. Walck†, C. Weinheimer§§, C. H. Wiebusch*, R. Wischniewski*, H. Wissing*, K. Woschnagg||, W. Wu§, G. Yodh§ & S. Young§

* Department of Physics, University of Wisconsin, Wisconsin, Madison 53706, USA

† Fysikum, Stockholm University, S-11385 Stockholm, Sweden

‡ Bartol Research Institute, University of Delaware, Newark, Delaware 19716, USA

§ Department of Physics and Astronomy, University of California, Irvine, California 92697, USA

|| Department of Physics, University of California, Berkeley, California 94720, USA

¶ Fachbereich 8 Physik, BUGH Wuppertal, D-42097 Wuppertal, Germany

Brussels Free University, Science Faculty CP230, Boulevard du Triomphe, B-1050 Brussels, Belgium

* DESY-Zeuthen, D-15735 Zeuthen, Germany

** Department of Radiation Sciences, Uppsala University, S-75121 Uppsala, Sweden

†† Department of Technology, Kalmar University, S-39129 Kalmar, Sweden

‡‡ Department of Physics and Astronomy, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA

§§ Institute of Physics, University of Mainz, Staudinger Weg 7, D-55099 Mainz, Germany

||| Institute for Nuclear and Particle Astrophysics, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA

Neutrinos are elementary particles that carry no electric charge and have little mass. As they interact only weakly with other particles, they can penetrate enormous amounts of matter, and therefore have the potential to directly convey astrophysical information from the edge of the Universe and from deep inside the most cataclysmic high-energy regions¹. The neutrino's great penetrating power, however, also makes this particle difficult to detect. Underground detectors have observed low-energy neutrinos from the Sun and a nearby supernova², as well as neutrinos generated in the Earth's atmosphere. But the very low fluxes of high-energy neutrinos from cosmic sources can be observed only by much larger, expandable detectors in, for example, deep water^{3,4} or ice⁵. Here we report the detection of upwardly propagating

atmospheric neutrinos by the ice-based Antarctic muon and neutrino detector array (AMANDA). These results establish a technology with which to build a kilometre-scale neutrino observatory necessary for astrophysical observations¹.

High-energy neutrinos must be generated in the same astrophysical sources that produce high-energy cosmic rays¹. These sources are a matter of speculation, but are thought to reside in shocked or violent environments such as are found in supernova remnants, active galactic nuclei, and gamma-ray bursters. The interaction of any high-energy proton or nucleus with matter or radiation in the source will produce neutrinos, some of which will have line-of-sight trajectories to Earth. AMANDA detects neutrinos with energies above a few tens of GeV by observing the Čerenkov radiation from muons that are produced in neutrino–nucleon interactions in the ice surrounding the detector or in the bedrock below⁶. This Čerenkov light is detected by an array of photomultiplier tubes (Fig. 1), which are buried deep in the ice in order to minimize the downward flux of muons produced in cosmic-ray interactions in the atmosphere. These muons constitute the main background for AMANDA. To ensure that the detected muons are produced by a neutrino, we use the Earth as a filter and look for upwardly propagating muons that perforce must have been produced by a neutrino that passed through the Earth. From the relative arrival times of the Čerenkov photons, measured with a precision of a few nanoseconds, we can reconstruct the track of the muon. The direction of the neutrino and muon are collinear within an angle $\theta_{\nu-\mu} \approx 1.5/\sqrt{E_{\nu}}$ degrees, where E_{ν} is measured in TeV, thus enabling us to search for point sources of high energy neutrinos.

Upwardly propagating atmospheric neutrinos are a well understood source that can be used to verify the detection technique. The results reported here are from analyses of experimental data acquired in 138 days of net operating time during the Antarctic winter of 1997. At that time the detector consisted of 302 optical modules deployed on ten strings at depths of between 1,500 m and 2,000 m (Fig. 1). The instrumented volume is a cylinder of approximately 120 m in diameter and 500 m in height. An optical module consists of an 8-inch photomultiplier tube housed in a glass

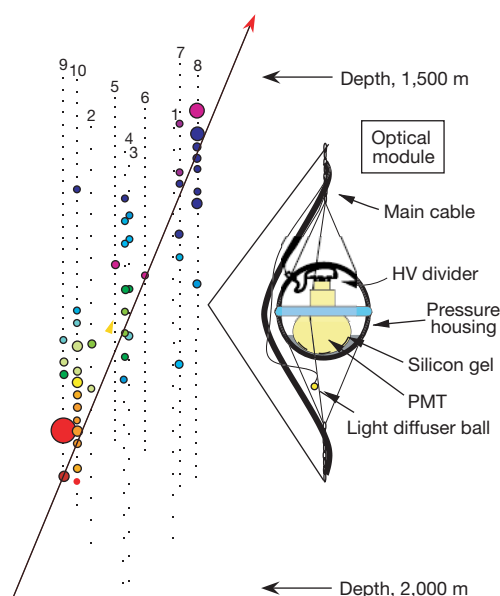


Figure 1 The AMANDA-B10 detector and a schematic diagram of an optical module. Each dot represents an optical module. The modules are separated by 20 m on the inner strings (1 to 4), and by 10 m on the outer strings (5 to 10). The coloured circles show pulses from the photomultipliers for a particular event; the sizes of the circles indicate the amplitudes of the pulses and the colours correspond to the time of a photon's arrival. Earlier times are in red and later ones in blue. The arrow indicates the reconstructed track of the upwardly propagating muon.

pressure vessel. A cable provides the high voltage and transmits the anode current signals to the data acquisition electronics at the surface. Figure 1 also shows a representative event that has satisfied the selection criteria for an upwardly moving muon. The effective detection area for muons varies from about 3,000 m² at 100 GeV to about 4×10^4 m² for the higher-energy muons (≥ 100 TeV) that would be produced by neutrinos coming from, for example, the same cosmic sources that produce gamma-ray bursts⁷.

Because a knowledge of the optical properties of the ice is essential for track reconstruction, these have been studied extensively^{8,9}. The absorption length of blue and ultraviolet light (the relevant wavelengths for our purposes) varies between 85 m and 225 m, depending on depth. The effective scattering length, which combines the mean free path λ with the average scattering angle θ through $\lambda/(1 - \langle \cos\theta \rangle)$, varies from 15 m to 40 m. In order to reconstruct the muon tracks we use a maximum-likelihood method, which incorporates the scattering and absorption of photons as determined from calibration measurements. A Bayesian formulation of the likelihood takes into account the much larger rate of downward muons relative to the upward signal and is particularly effective in decreasing the chance for a downward

muon to be mis-reconstructed as upward. (See refs 6 and 10–13 for more information on optical properties of ice, calibration, and analysis techniques.)

Certain types of events that might appear to be upwardly propagating muons must be considered and eliminated. Rare cases, such as muons that undergo catastrophic energy loss through bremsstrahlung, or that are coincident with other muons, must be investigated. To this end, a series of requirements or quality criteria, based on the characteristic time and spatial pattern of photons associated with a muon track and the response of our detector, are applied to all events that, in a first assessment, appear to be upwardly moving muons. For example, an event that has a large number of optical modules hit by prompt (that is, unscattered) photons, has a high quality. By making these requirements (or 'cuts') increasingly selective, we eliminate correspondingly more of the background of false upward events while still retaining a significant fraction of the true upwardly moving muons. Because there is a large space within which the parameters defining these cuts can be optimized, two different and independent analyses of the same set of data have been undertaken. These analyses yielded comparable numbers of upwardly propagating muons (153 in analysis A, 188 in analysis B). Comparison of these results with their respective Monte Carlo simulations shows that they are consistent with each other in terms of the number of events, the number of events in common and, as discussed below, the expected properties of atmospheric neutrinos.

In Fig. 2a, the number of experimental events is compared to simulations of background and signal as a function of the (identical) quality requirements placed on the three types of events: experimental data, simulated upwardly moving muons from atmospheric neutrinos, and a simulated background of downwardly moving cosmic-ray muons. For simplicity in presentation, the levels of the individual cuts have been combined into a single parameter representing the overall event quality. Figure 2b shows ratios of the quantities plotted above. As the quality level is increased, the ratios of simulated background to experimental data, and of experimental data to simulated signal, both continue their rapid decrease, the former toward zero and the latter toward 0.7. Over the

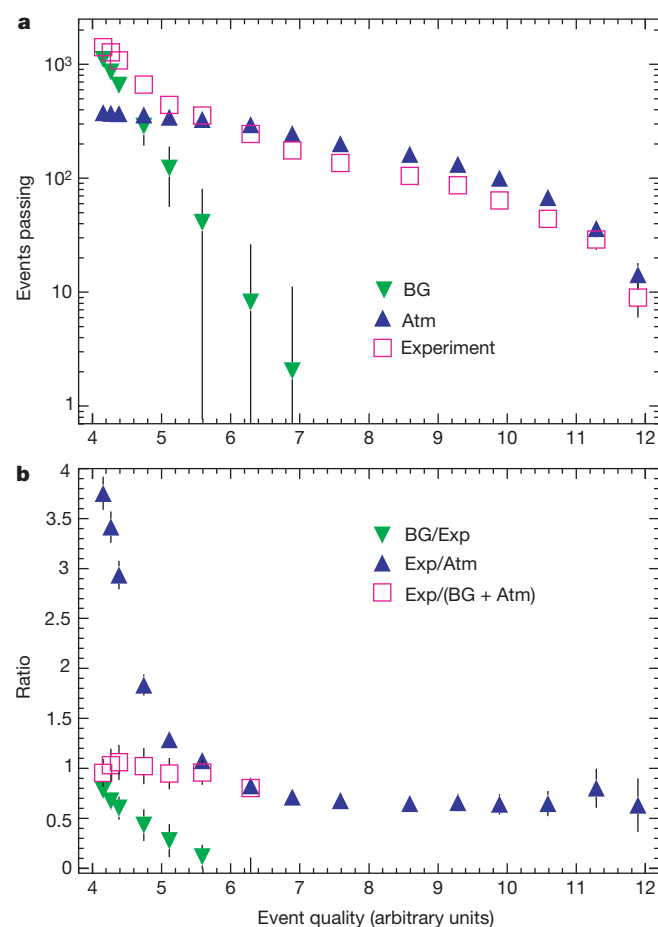


Figure 2 Experimental data confront expectations. The numbers of reconstructed upwardly moving muon events for the experimental data (Exp) from analysis A are compared to simulations of background cosmic-ray muons (BG) and simulations of atmospheric neutrinos (Atm) as a function of 'event quality', a variable indicating the combined severity of the cuts designed to enhance the signal. The comparison begins at a quality level of 4. Cuts were made on a number of parameters including the reconstructed zenith angle (>100 degrees), maximum likelihood of the reconstruction, topological distributions of the detected photons, and the number of optical modules recording unscattered photons. The optimum levels of the cuts were determined by comparing the relative rejection rates for Monte Carlo simulated neutrino events and background events. **b**, Ratios of the quantities shown in **a**.

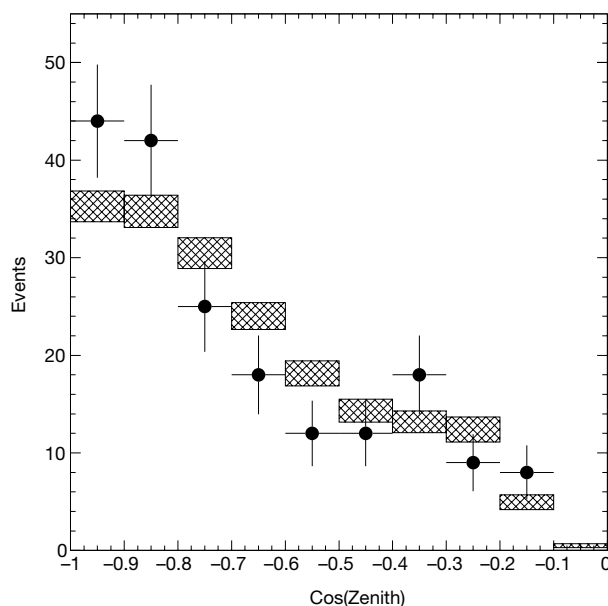


Figure 3 Reconstructed zenith angle distribution. The data points are experimental data (from analysis B) and the shaded boxes are a simulation of atmospheric neutrino events, the widths of the boxes indicating the error bars. The overall normalization of the simulation has been adjusted to fit the data. The possible effects of neutrino oscillations on the flux and its zenith angle dependence estimated from the Super-Kamiokande measurements²⁰, are expected to be small in our energy range.

same range, the ratio of experimental data to the simulated sum of background and signal remains nearly constant. We conclude that the quality requirements have reduced the presence of wrongly reconstructed downward muons in the experimental data to a negligible fraction of the signal and that the experimental data behave in the same way as the simulated atmospheric neutrino signal for events that pass the stringent cuts. The estimated uncertainty on the number of events predicted by the signal Monte Carlo simulation, which includes uncertainties in the high-energy atmospheric neutrino flux, the *in situ* sensitivity of the optical modules, and the precise optical properties of the ice, is +40%/–50%. The observed ratio of experiment to simulation (0.7) and the expectation (1.0) therefore agree within the estimated uncertainties.

The shape of the zenith-angle distribution of the 188 events from analysis B is compared to a simulation of the atmospheric neutrino signal in Fig. 3, where the absolute Monte Carlo rate has been normalized to the experimental rate. The variation of the measured rate with zenith angle is reproduced by the simulation to within the statistical uncertainty. We note that the tall geometry of the detector favours the more vertical muons. The arrival directions of the upwardly moving muons observed in both analyses are shown in Fig. 4. A statistical analysis indicates no evidence for point sources in these samples. The agreement between experiment and simulation of atmospheric neutrino signal, as demonstrated in Figs 2 and 3, taken together with comparisons for a number of other variables (to be published elsewhere) leads us to conclude that the upcoming muon events observed by AMANDA are produced mainly by atmospheric neutrinos with energies of about 50 GeV to a few TeV. The background in this event sample is estimated to be $15 \pm 7\%$ events, and is due to misreconstructions.

From the consistency of the selected event sample with muons generated by atmospheric neutrinos, and in particular the absence of an excess of high-energy events with a large number of optical modules that had been hit, we can determine an upper limit on a diffuse extraterrestrial neutrino flux. Assuming a hard E^{-2} spectrum characteristic of shockwave acceleration, we expect to reach a sensitivity of order $dN/dE_\nu = 10^{-6} E_\nu^{-2} \text{ cm}^{-2} \text{ s}^{-1} \text{ sr}^{-1} \text{ GeV}^{-1}$. This value is low enough to be in the range where a number of models^{14–19} predict fluxes, a few of which are larger^{15,16}. Most recent estimates are smaller^{17–19}. The present level of sensitivity and the prospects for improving it through longer exposure times and better determination of muon energy illustrate the ability of large-area detectors to test theoretical models that assume the hadronic origin of TeV photons from active galaxies—models which would be difficult to confirm or exclude without the ability to observe high-energy neutrinos.

Searches for neutrinos from gamma-ray bursts, for magnetic monopoles, supernova collapses and for a cold dark matter signal from the centre of the Earth are also in progress and, with only 138 days of data, yield limits comparable to or better than those from

smaller underground neutrino detectors that have operated for a much longer period (see refs 10–13).

From 1997 to 1999 an additional nine strings were added in a concentric cylinder around AMANDA-B10. This larger detector, called AMANDA-II, consists of 677 optical modules and has an improved acceptance for muons over a larger angular interval. Data are being taken now with the larger array. Yet the fluxes of very high energy neutrinos predicted by theoretical models¹⁴ or derived from the observed flux of ultra high energy cosmic rays¹⁹ are sufficiently low that a neutrino detector having an effective area up to a square kilometre is required for their observation and study^{1,14}. Plans are therefore being made for a much larger detector, IceCube, consisting of 4,800 photomultipliers to be deployed on 80 strings. This proposed neutrino telescope would have an effective area of about 1 km^2 , an energy threshold near 100 GeV and a pointing accuracy for muons of better than one degree for high-energy events. In conclusion, the observation of neutrinos by a neutrino telescope deep in the Antarctic ice cap, a goal that was once thought difficult if not impossible, represents an important step toward establishing the field of high-energy neutrino astronomy first envisioned over 40 years ago. □

Received 15 September 2000; accepted 25 January 2001.

- Gaisser, T. K., Halzen, F. & Stanev, T. Particle physics with high-energy neutrinos. *Phys. Rep.* **258**, 173–236 (1995).
- Totsuka, Y. Neutrino astronomy. *Rep. Prog. Phys.* **55**, 377–430 (1992).
- Roberts, A. The birth of high-energy neutrino astronomy: A personal history of the DUMAND project. *Rev. Mod. Phys.* **64**, 259–312 (1992).
- Balkanov, V. A. *et al.* An upper limit on the diffuse flux of high energy neutrinos obtained with the Baikal detector NT-96. *Astropart. Phys.* **14**, 61–67 (2000).
- Lowder, D. M. *et al.* Observation of muons using the polar ice as a Čerenkov detector. *Nature* **353**, 331–333 (1991).
- Andres, E. *et al.* The AMANDA neutrino telescope: Principle of operation and first results. *Astropart. Phys.* **13**, 1–20 (2000).
- Waxman, E. & Bahcall, J. N. High-energy neutrinos from cosmological gamma ray burst fireballs. *Phys. Rev. Lett.* **78**, 2292–2295 (1997).
- Askjober, P. *et al.* Optical properties of the south pole ice at depths between 0.8 and 1 km. *Science* **267**, 1147–1150 (1995).
- Price, B. P. Implications of optical properties of ocean, lake, and ice for ultrahigh-energy neutrino detection. *Appl. Opt.* **36**, 1965–1975 (1997).
- Wischniewski, R. *et al.* in *Proc. 26th Int. Cosmic Ray Conf., Salt Lake City Vol. 2* (eds Kieda, D., Salamon, M. & Dingus, B.) 229–232 (1999).
- Dalberg, E. *et al.* in *Proc. 26th Int. Cosmic Ray Conf., Salt Lake City Vol. 2* (eds Kieda, D., Salamon, M. & Dingus, B.) 348–351 (1999).
- Bay, R. *et al.* in *Proc. 26th Int. Cosmic Ray Conf., Salt Lake City Vol. 2* (eds Kieda, D., Salamon, M. & Dingus, B.) 225–228 (1999).
- Niessen, P. *et al.* in *Proc. 26th Int. Cosmic Ray Conf., Salt Lake City Vol. 2* (eds Kieda, D., Salamon, M. & Dingus, B.) 344–347 (1999).
- Learned, J. G. & Mannheim, K. High-energy neutrino astrophysics. *Ann. Rev. Nucl. Sci.* **50**, 679–749 (2000).
- Szabo, A. P. & Protheroe, R. J. Implications of particle acceleration in active galactic nuclei for cosmic rays and high-energy neutrino astronomy. *Astropart. Phys.* **2**, 375–392 (1994).
- Stecker, F. W. & Salamon, M. H. High-energy neutrinos from quasars. *Space Sci. Rev.* **75**, 341–355 (1996).
- Nellen, L., Mannheim, K. & Biermann, P. L. Neutrino production through hadronic cascades in AGN accretion disks. *Phys. Rev. D* **47**, 5270–5274 (1993).
- Mannheim, K., Protheroe, R. J. & Rachen, J. P. Cosmic ray bound for models of extragalactic neutrino production. *Phys. Rev. D* **63**, 023003–1–023003–16 (2001).
- Waxman, E. & Bahcall, J. N. High energy neutrinos from astrophysical sources. An upper bound. *Phys. Rev. D* **59**, 023002–1–023002–8 (1999).
- Fukuda, Y. *et al.* (Super-Kamiokande Collaboration) Measurement of the flux and zenith-angle distribution of upward through-going muons by Super-Kamiokande. *Phys. Rev. Lett.* **82**, 2644–2648 (1999).

Acknowledgements

This research was supported by the following agencies: US National Science Foundation, Office of Polar Programs; US National Science Foundation, Physics Division; University of Wisconsin Alumni Research Foundation; US Department of Energy; Swedish Natural Science Research Council; Swedish Polar Research Secretariat; Knut and Alice Wallenberg Foundation, Sweden; German Ministry for Education and Research; US National Energy Research Scientific Computing Center (supported by the Office of Energy Research of the US Department of Energy); UC-Irvine AENEAS Supercomputer Facility; Deutsche Forschungsgemeinschaft (DFG). D.F.C. acknowledges the support of the NSF CAREER programme and C.P.d.I.H. acknowledges support from the European Union 4th Framework of Training and Mobility of Researchers.

Correspondence and requests for materials should be addressed to F.H. (e-mail: halzen@pheno.physics.wisc.edu).

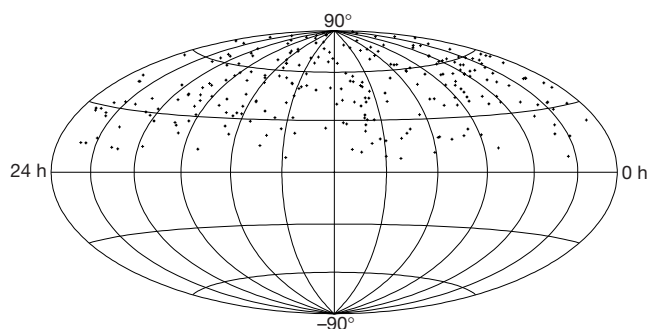


Figure 4 Distribution in declination and right ascension of the upwardly propagating events on the sky. The 263 events shown here are taken from the upward muons contained in both analysis A and analysis B. The median difference between the true and the reconstructed muon angles is about 3 to 4 degrees.

Stable ultrahigh-density magneto-optical recordings using introduced linear defects

L. Krusin-Elbaum^{*†}, T. Shibauchi^{*†‡}, B. Argyle^{*}, L. Gignac^{*} & D. Weller^{§||}

^{*} IBM T. J. Watson Research Center, Yorktown Heights, New York 10598, USA

[‡] MST-STC, Los Alamos National Laboratory, MS-K763, Los Alamos, New Mexico 87545, USA

[§] IBM Almaden Research Center, San Jose, California 95120, USA

[†] These authors contributed equally to this work

The stability of data bits in magnetic recording media^{1,2} at ultrahigh densities is compromised by the thermal ‘flips’—magnetic spin reversals—of nano-sized spin domains³, which erase the stored information. Media that are magnetized perpendicular to the plane of the film, such as ultrathin cobalt films or multilayered structures^{4,5}, are more stable against thermal self-erasure^{2,6} than conventional memory devices. In this context, magneto-optical memories seem particularly promising for ultrahigh-density recording on portable disks, and bit densities of ~ 100 Gbit inch⁻² (ref. 7) have been demonstrated using recent advances in the bit writing and reading techniques^{7–11}. But the roughness and mobility of the magnetic domain walls^{12,13} prevents closer packing of the magnetic bits, and therefore presents a challenge to reaching even higher bit densities. Here we report that the strain imposed by a linear defect in a magnetic thin film can smooth rough domain walls over regions hundreds of micrometres in size, and halt their motion. A scaling analysis of this process, based on the generic physics of disorder-controlled elastic lines^{14–17}, points to a simple way by which magnetic media might be prepared that can store data at densities in excess of 1 Tbit inch⁻².

Increasing information storage to densities past 100 Gbit inch⁻² may evolve through extensions of current magnetic recording technologies (to patterned media⁶, for example). But such increases in storage density might be achieved by using other techniques such as holography (via interference patterns produced by two intersecting laser beams)¹⁸, or micromachined nano-cantilever arrays¹⁹, or—to satisfy a relentless demand for portability—using scanning localized subwavelength ($< \lambda/40$) (near-field) optical probes⁹ that can imprint and resolve images in magneto-optic media⁷ of the order of the probe size. The bit-writing with local probes may be thermally assisted by a current²⁰ or a laser beam that raises local temperature to the vicinity of the Curie temperature T_C , resulting in the formation of a reversed domain with a rough wall. To realize smooth and (for precise positioning) stable (‘noiseless’) domain walls that can be implemented without, for example, nanoscale patterning, we suggest a conceptual identification of walls with elastic lines¹⁶ and seek to use linear defects inducing a directional long-range strain field in ultrathin Pt/Co/Pt structures—long considered excellent candidates for high-density recording at blue-range wavelengths¹¹. In such trilayers (Fig. 1), large uniaxial perpendicular anisotropy is sustained by the interface contribution proportional to $1/d_{Co}$ up to Co thickness $d_{Co} \approx 1.5$ nm; beyond this thickness the magnetization switches from out-of-plane to in-plane⁵.

Figure 2a shows two typical up-domains (see legend for imaging details²¹), nucleated owing to a locally suppressed coercive field³ H_c in a 0.7-nm Co film sandwiched between 3-nm (top) and 2-nm (bottom) Pt layers, prepared using standard deposition parameters²² and without linear defects. The domains are round, with

(undirected) domain walls (DWs) that are as rugged as expected¹². We note that patterned nucleation sites, as in Fig. 2b–d, do not reduce either DW roughness or velocity. As magnetic field is increased beyond $H_c = 750$ Oe (inset in Fig. 3), the outward motion of DWs becomes increasingly swift (estimated from displacement of a small segment of the wall during 500-ms field pulses, the wall velocity is more than $180 \mu\text{m s}^{-1}$ at $H = 854$ Oe).

To control the rugged DW structure we will use a recently found connection between the DW behaviour in thin films¹³ and that of elastic lines^{14–16}. One prominent example of directed elastic lines are wandering vortex lines (quantized magnetic flux lines maintained by a swirling tube of current) in high-temperature superconductors¹⁶ that can be strongly localized by interaction with columnar (linear) defects¹⁷. A powerful arsenal of ideas can now be engaged to understand how a linear defect potential may localize and reduce the roughness of DWs, and force them to accommodate to the defect shape. We introduce a line defect that delivers a three-punch action. Through magneto-elastic coupling, (1) it gives the wall a preferred direction¹⁴ (along the defect), (2) it increases its elasticity, reducing the DW roughness as it negotiates the random landscape, and (3) it acts to reduce wall velocity to a nearly full stop in fields greater than the coercive field of unmodified film. The defect is installed during the Co deposition by imposing an anisotropic tension (clamping) on the substrate and its subsequent release²³. The resulting Co ‘fold’ (Fig. 2g, h) introduces a y -axis-invariant long-range strain field $\epsilon(x)$.

A DW driven by the magnetic field toward such a linear defect (Fig. 2e, f) presents a structural contrast with the DWs in the unmodified film in Fig. 2a. Even at large distances ($\sim 300 \mu\text{m}$) away from the defect, the DW conforms on the average to the defect line along y . It becomes progressively smoother (and straighter) as it approaches the defect. It also rapidly decelerates. The deceleration and near-standstill of the wall depends on the proximity to the defect, as represented by the spatial progression of the velocity-versus-field (force) response in Fig. 3. The v - H curves are highly nonlinear, as has been reported recently in Co films¹³ where the disorder landscape is formed, for example, by atomic-scale imperfections at the film–substrate interface. Such nonlinear response in the limit of vanishing driving force is a signature of glassy (creep) dynamics²⁴, well established for the elastic vortex lines in a superconductor through measurements of the voltage-versus-current (V - I) characteristics¹⁶. Above H_{crit} (obtained from the v - H curves by the usual velocity cut-off criterion, here chosen at $v = 0.14 \mu\text{m s}^{-1}$) the driving field exceeds the ‘pinning’ force¹⁶ and the DW response becomes linear and faster. We note that DW velocity, even far away from a line defect, is orders of magnitude lower than in unmodified films. Figure 3 shows a field-forward advance of v - H curves to higher fields, and an enhanced H_{crit} (often referred to as a ‘propagation field’¹³) on the approach to the line defect. An effective potential well that localizes the wall is formed by the driving field pushing the wall and the line-defect that acts against this push. It resembles columnar defects in a superconductor, where the critical current J_{crit} is enhanced¹⁷. H_{crit} correlates with the long-range repelling force field $H(x)$ exerted on the DWs by the line-defect, whose spatial extent is mapped in Fig. 4a. The shape of $H(x)$ —a ridge along y —is either extracted directly from Kerr

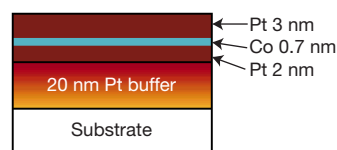


Figure 1 A sketch of an ultrathin Pt/Co/Pt trilayer stack explored in this study. Stacks were electron-beam evaporated at 190 °C on glass or SiN_x/Si substrates with a 20-nm (111) textured Pt buffer layers, as described in ref. 22.

^{||} Present address: Seagate Technology, Pittsburgh, Pennsylvania 15203, USA.

images taken with increasing H after a fixed propagation time t (main panel), or more quantitatively from the averaged DW positions $\langle x \rangle$, versus time at all fields. As illustrated in the inset, it takes a higher field H to get closer to the line-defect; but at any H , after $t = 1,800$ s, two DWs—one approaching the ridge from the left and another from the right—become effectively stationary.

Within the elastic description of a DW, the relevant scale is a collective pinning²⁴ length $L_p = (\epsilon_{\text{el}}^2 \xi^2 / \Delta)^{1/3}$, where ξ and Δ are characteristic size and strength of underlying random disorder, and ϵ_{el} is the wall energy. At lengths $L > L_p$, the wall will elastically adjust to the random landscape to nestle in a local minimum energy configuration. The DW energy density (in addition to a uniform field term and ignoring a weak dipolar term) can be written as a sum of three: the exchange energy $\gamma_{\text{ex}}(x)$, the anisotropy energy $\gamma_{\text{an}}(x)$, and the magneto-elastic energy²⁵ $\gamma_{\text{m-el}}(x)$ coupling to the strain $\epsilon(x)$ generated by the line-defect,

$$\gamma_{\text{DW}} = \underbrace{A(d\theta/dx)^2}_{\gamma_{\text{ex}}} - \underbrace{K'm_z^2}_{\gamma_{\text{an}}} - \underbrace{B\epsilon(x)m_z^2}_{\gamma_{\text{m-el}}} + H^2/8\pi \quad (1)$$

Here A is the exchange stiffness³, $K' = K - 2\pi M_s^2$ (where K is the anisotropy constant and M_s is Pt-polarization-enhanced saturation magnetization of Co), and B is related to Young's modulus of the film. Magnetization $\mathbf{M}$ rotates from 'up' to 'down' within the wall thickness δ (we estimate at ~ 3 nm) and the wall (in the simplest form) is of Bloch type¹³, where the azimuthal angle $\phi = 0^\circ$ and the rotation is parameterized by an angle θ between the z axis and $\mathbf{M}$ (m_z is the direction cosine along z). Minimizing $\int \gamma_{\text{DW}} dx$ leads to DW energy³ $\epsilon_{\text{DW}} = 4\sqrt{AK'_{\text{eff}}}$, where $K'_{\text{eff}} = K' + B\epsilon(x)$. The total wall

energy (per unit area) in a magnetic field H is $\epsilon = \epsilon_{\text{DW}} - 2M_s Hx$. From the stability condition $d\epsilon/dx = 0$ we obtain a nonlinear differential equation for $\epsilon(x)$: $H(x) = AM_s^{-1} B(d\epsilon/dx)(AK'_{\text{eff}})^{-1/2}$, which we solve numerically using the stationary $H(x)$ mapped in Fig. 4a. The result is a factor of ~ 2 enhanced effective anisotropy K'_{eff}

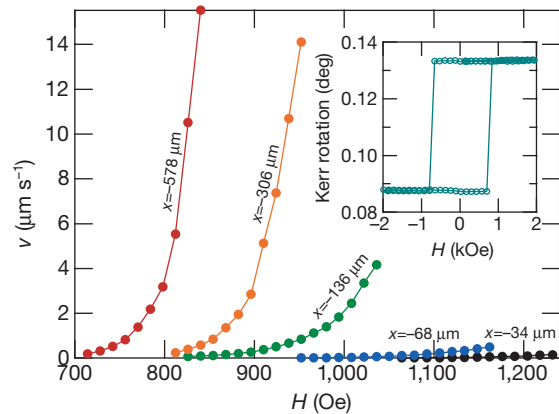


Figure 3 Domain wall velocity as a function of applied magnetic field for different values of x . Far away ($-578 \mu\text{m}$) from the line-defect 'ridge' at $x = 0$, the velocity takes off rapidly with field near H_{crit} . Well below H_{crit} , the nonlinear v - H curves follow¹³ $v(H) \propto \exp[-U_p/k_B T](H_{\text{crit}}/H)^\mu$, where U_p is the collective pinning energy scale, k_B is the Boltzmann constant, T is temperature, and $\mu = 1/4$ is the glassy dynamical exponent related to the roughness exponent; see ref. 14. H_{crit} grows on the approach to the line defect (see Fig. 4a). Inset, a square Kerr rotation (magnetization) hysteresis characteristic of perpendicular anisotropy of our film in Fig. 2a.

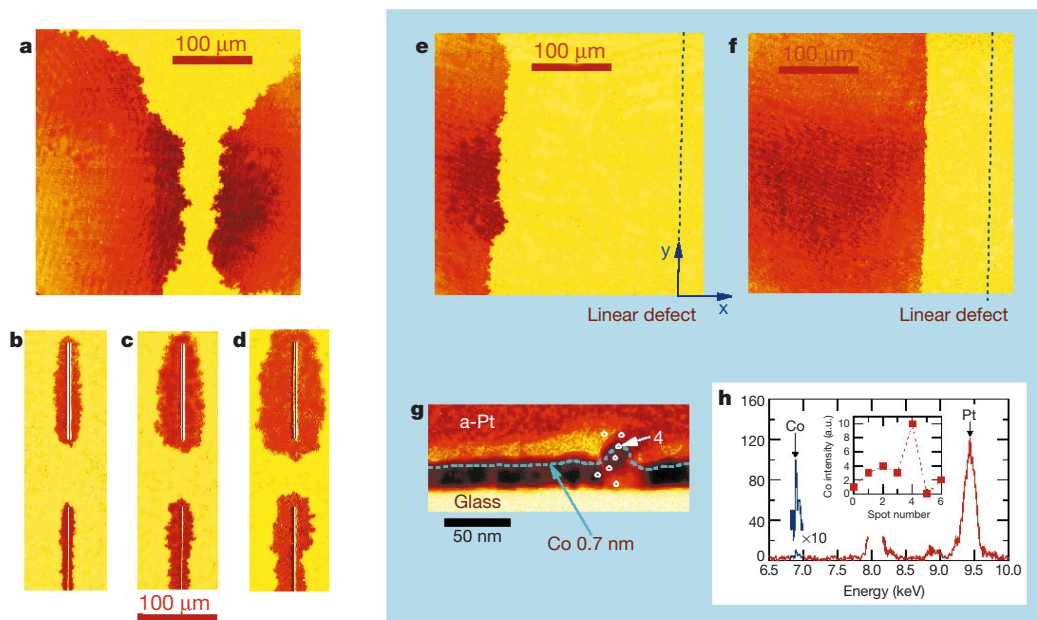


Figure 2 Effect of a linear defect on domain walls in Pt/Co/Pt films. Polar magneto-optic Kerr microscopy³ images were obtained using a 3-W argon laser illumination ($\lambda = 514.5$ nm and 488.0 nm) to enhance phase contrast with a laser-beam tandem-dithering technique²¹ to eliminate the effects of laser coherence. We first saturated magnetization with a negative z -axis magnetic field and then quenched the field to zero. After a first positive square field pulse, an up-domain was nucleated and its motion was imaged at short time intervals in fields up to ~ 2 kOe. **a**, Typical room-temperature image of domains in a film without line defects after a 854-Oe magnetic field applied for 1 s was removed. The walls show roughness on length scales much larger than the grain size (~ 20 – 30 nm) of the Pt buffer. The rapid motion of domain walls (DWs) is illustrated in snapshots in field $H = 616$ Oe after **b** 1 s, **c**, 4 s and **d**, 8 s of DWs that were seeded with

elongated defects installed with a 30-keV Ga^+ FIB (focused ion beam). DW roughness is independent of the seed-defect width (150 nm for bottom and 1 μm for top domains). **e, f**, A remarkable reduction in the domain wall roughness and speed is illustrated when a line-defect (dashed line along y axis) is introduced. The images here were taken at $H = 924$ Oe ($> H_c = 750$ Oe in **a**) at **e**, 40 s and **f**, 1,800 s. The domain wall becomes directed along the defect. A cross-sectional electron transmission micrograph (**g**) of an ~ 30 -nm-thin FIBed section—normal to the line-defect—shows an asymmetric 'fold' of the trilayer stack. An amorphous Pt (a-Pt) cap was deposited to protect the trilayer during the FIB process. The ~ 10 -nm elevation of Co in the 'fold' is traced by the X-ray microprobe (**h**) at spots such as are marked by six white circles in **g** (the bottom is assigned '0'). Co intensity versus spot number is plotted in the inset of **h**.

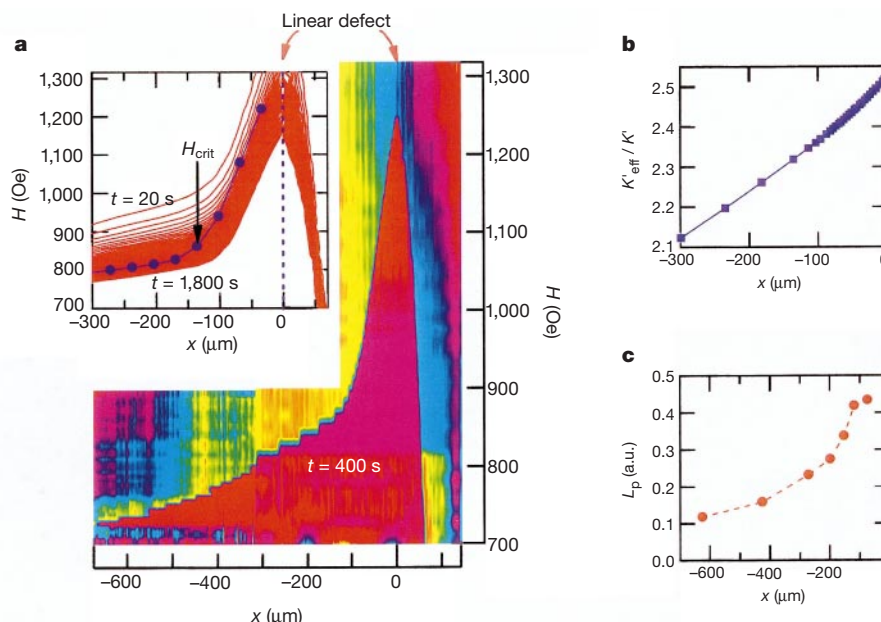


Figure 4 A force field $H(x)$ generated by a line defect along y repels a magnetic-field-driven domain wall. **a**, A fixed y -cut from a map of the defect ridge obtained from combining domain images for different fields after 400 s. To display the long range ($x = -680$ to $+140 \mu\text{m}$) of the force field of the line defect, we overlapped two series of Kerr observations ($H = 700$ to $1,316$ Oe): one containing a line defect at $x = 0$ (as in the inset) and another shifted to the left by $400 \mu\text{m}$. The ridge is mapped by imaging two domains approaching it from both sides. Contours (at 20-s intervals) of the defect field

(inset) that were obtained by extracting DW positions at different times and fields show how two DWs on two sides of the ridge become stationary when the defect field balances the driving field, defining the ridge's outline. The propagation field H_{crit} (solid blue dots), obtained from v - H curves in Fig. 3 is enhanced by the line defect. The line defect effectively enhances **(b)** the anisotropy constant K_{eff} and elongates **(c)** the scaling length L_p .

(Fig. 4b) and, consequently, an enhanced DW elastic energy (per unit length)¹³ $\epsilon_{\text{el}} = 4\sqrt{AK_{\text{eff}}^2 d_{\text{Co}}}$. Thus, the enhanced elasticity of DW is in part responsible for the longer L_p . The longer L_p is a key to the reduced DW roughness. This can be seen from the analysis of transverse displacements of the domain wall segments by computing from the DW images a (line shape) correlator¹⁴ $x_t = \sqrt{[x(y) - x(y+L)]^2}$, predicted for $L < L_p$ to scale as¹⁶ $x_t \propto (L/L_p)^{3/2}$. We find that the characteristic length L_p is enhanced in the vicinity of the line-defect by a factor of ~ 5 (Fig. 4c). This implies a reduction of roughness, as measured by x_t , by a factor of ~ 10 on the shortest length scales and hence a potential increase in the areal density by nearly two orders of magnitude. In its proximity, the line defect wins over randomness so that on sufficiently long distances along its length DWs become essentially 'flat' (that is, finite x_t for $L \rightarrow \infty$)¹⁴ in fields up to $\sim 2H_c$. The wall-smoothing at higher fields and at speeds of current recording technology ($H \approx 3H_c \sim 1 \text{ Gbit s}^{-1}$) deserves further study for practical implementations of this effect. The long-range strain associated with a line defect presents an efficient, and relatively simple and cost-effective control of domain walls, with only a few such defects needed to stabilize large areas of ultrathin films. □

Received 14 June; accepted 28 December 2000.

- Daniel, E. D., Mee, C. D. & Clark, M. H. (eds) *Magnetic Recording: The First Hundred Years* (IEEE Press, New York, 1998).
- Thompson, D. A. & Best, J. S. The future of magnetic data storage technology. *IBM J. Res. Develop.* **44**, 311–322 (2000).
- Hubert, A. & Schäfer, R. *Magnetic Domains* (Springer, New York, 1998).
- Kirby, R. D., Shen, J. X., Hardy, R. J. & Sellmyer, D. J. Magnetization reversals in nanoscale films with perpendicular anisotropy. *Phys. Rev. B* **65**, R10810–R10813 (1994).
- Allenspach, R., Stampanoni, M. & Bischof, A. Magnetic domains in thin epitaxial Co/Au(111) films. *Phys. Rev. Lett.* **65**, 3344–3347 (1990).
- Weller, D. & Moser, A. Thermal effect limits in ultrahigh density magnetic recording. *IEEE Trans. Magn.* **35**, 4423–4439 (1999).
- Partovi, A. Optical near-field aperture storage technique for high density, high performance data

- storage applications. *SPIE* **3864**, 352–354 (1999).
- Awano, H. & Ohta, N. Magneto-optical recording technology toward 100 Gb/in². *IEEE J. Select. Topics Quantum Electron.* **4**, 815–820 (1998).
- Betzig, E., Trautman, J. K., Harris, T. D., Weiner, J. S. & Kostelak, R. L. Breaking the diffraction barrier: optical microscopy on a nanometer scale. *Science* **251**, 1468–1470 (1991).
- Partovi, A. et al. High-power laser light source for near-field optics and its application to high-density optical data storage. *Appl. Phys. Lett.* **75**, 1515–1517 (1999).
- Kaneko, M. Magnetic multilayer films for high density magneto-optical recording. *J. Magn. Magn. Mater.* **148**, 351–356 (1995).
- Han, S. K., Yu, S.-C. & Rao, K. V. Domain wall jaggedness induced by the random anisotropy orientation in magneto-optic materials: A computer simulation study. *J. Appl. Phys.* **79**, 4260–4264 (1996).
- Lemerle, S. et al. Domain wall creep in an Ising ultrathin magnetic film. *Phys. Rev. Lett.* **80**, 849–852 (1998).
- Halpin-Healy, T. & Zhang, Y.-C. Kinetic roughening phenomena, stochastic growth, directed polymers and all that. *Phys. Rep.* **254**, 215–414 (1995).
- Hwa, T. From vortices to genomics. *Nature* **399**, 17–18 (1999).
- Blatter, G., Feigel'man, M. V., Geshkenbein, V. B., Larkin, A. I. & Vinokur, V. M. Vortices in high-temperature superconductors. *Rev. Mod. Phys.* **66**, 1125–1388 (1994).
- Krusin-Elbaum, L. et al. Superconductivity enhanced by Hg fission. *Nature* **389**, 243–244 (1997).
- Heanue, J. F., Bashaw, M. C. & Hasselink, L. Volume holographic storage and retrieval of digital data. *Science* **265**, 749–752 (1994).
- Binnig, G. K., Rohrer, H. & Vettiger, P. Mass-storage applications of local probe arrays. US Patent 5,835,477 (1998).
- Nakamura, J., Miyamoto, M., Hosaka, S. & Koyanagi, H. High-density thermomagnetic recording method using a scanning tunneling microscope. *J. Appl. Phys.* **77**, 779–781 (1995).
- Argyle, B. E. & McCord, J. G. New laser illumination method for Kerr microscopy. *J. Appl. Phys.* **87**, 6487–6489 (2000).
- Weller, D. et al. Ion induced magnetization reorientation in Co/Pt multilayers for patterned media. *J. Appl. Phys.* **87**, 5768–5770 (2000).
- Deng, H., Jarratt, J. D., Minor, M. K. & Barnard, J. A. Artificially controlled stress anisotropy and magnetic properties of Fe/Ta thin films. *J. Appl. Phys.* **81**, 4510–4512 (1997).
- Feigel'man, M. V., Geshkenbein, V. B., Larkin, A. I. & Vinokur, V. M. Theory of collective flux creep. *Phys. Rev. Lett.* **63**, 2303–2306 (1989).
- Chuang, D. S., Ballentine, C. A. & O'Handley, R. C. Surface and step magnetic anisotropy. *Phys. Rev. B* **49**, 15084–15095 (1994).

Acknowledgements

We thank J. Slonczewski, V. Vinokur, G. Blatter, and G. Zimanyi for useful discussions.

Correspondence and requests for materials should be addressed to L.K.E. (e-mail: krusin@us.ibm.com).

Giant lateral electrostriction in ferroelectric liquid-crystalline elastomers

W. Lehmann*, H. Skupin*, C. Tolksdorf†, E. Gebhard†, R. Zentel†, P. Krüger*, M. Lösche* & F. Kremer*

* Institut für experimentelle Physik I, Universität Leipzig, Linnéstraße 5, 04103 Leipzig, Germany

† Institut für organische Chemie, Universität Mainz, Duesbergweg 10-14, 55099 Mainz, Germany

Mechanisms for converting electrical energy into mechanical energy are essential for the design of nanoscale transducers, sensors, actuators, motors, pumps, artificial muscles, and medical microrobots. Nanometre-scale actuation has to date been mainly achieved by using the (linear) piezoelectric effect in certain classes of crystals (for example, quartz), and 'smart' ceramics such as lead zirconate titanate. But the strains achievable in these materials are small—less than 0.1 per cent—so several alternative materials and approaches have been considered. These include grafted polyglutamates¹ (which have a performance comparable to quartz), silicone elastomers² (passive material—the constriction

results from the Coulomb attraction of the capacitor electrodes between which the material is sandwiched) and carbon nanotubes³ (which are slow). High and fast strains of up to 4 per cent within an electric field of 150 MV m^{-1} have been achieved by electrostriction (this means that the strain is proportional to the square of the applied electric field) in an electron-irradiated poly(vinylidene fluoride-trifluoroethylene) copolymer⁴. Here we report a material that shows a further increase in electrostriction by two orders of magnitude: ultrathin (less than 100 nanometres) ferroelectric liquid-crystalline elastomer films that exhibit 4 per cent strain at only 1.5 MV m^{-1} . This giant electrostriction was obtained by combining the properties of ferroelectric liquid crystals with those of a polymer network. We expect that these results, which can be completely understood on a molecular level, will open new perspectives for applications.

Liquid-crystal molecules, the so-called mesogens, may be easily reorientated under the influence of electric fields when sandwiched between transparent ITO (indium tin oxide) electrodes, thus changing the transmittance of light between crossed polarizers; this is the basis of operation of the liquid-crystal display. So why not make use of this electrically induced motion for the construction of high-performance actuators? The problem is that (smectic) liquid crystals have a crystalline ordering only in one dimension, perpendicular to the layer planes. In contrast, in the other two dimensions—in the plane of the layers—they have a liquid-like structure. This means that they can neither sustain nor exert mechanical stress—

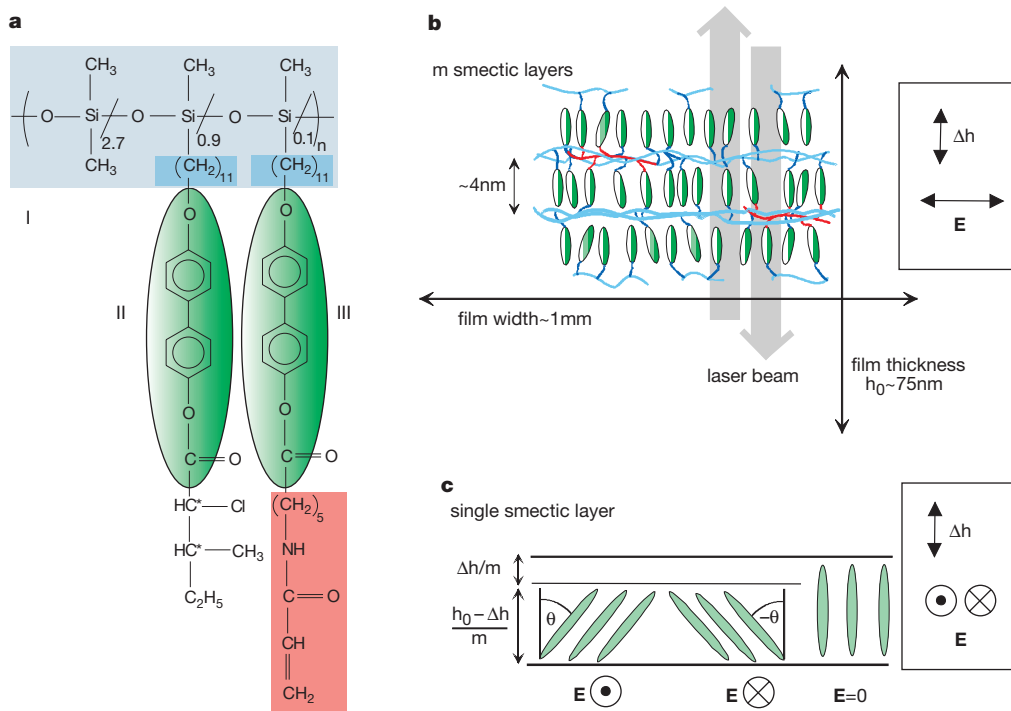


Figure 1 The electroclinic effect in ferroelectric liquid crystalline elastomers (FLCEs). **a**, Chemical structure of the sample¹³. Parts of the structure are shown coloured, and some parts are also indicated by roman numerals. I, blue; the polysiloxane backbone. II, green; the core of the chiral (*) mesogen (90% of the mesogenic side groups). III, green; the core of the crosslinkable mesogen (10% of the mesogenic side groups). Red; the crosslinkable end group of the mesogen. The mesogens are attached to the backbone via flexible alkyl spacers $(\text{CH}_2)_{11}$ —(marked in dark blue). The (weight averaged) degree of polymerization is $n \approx 108$. **b**, Scheme of the measurement geometry. The mesogenic side groups are depicted as green ellipsoids, while the alkyl spacers and the polysiloxane backbone are drawn as blue lines; the colours refer to **a**. Each of the m smectic layers in this sample (of which only three are shown) has a thickness of $\sim 4 \text{ nm}$ in the S_A^* phase. The layers are arranged parallel to the plane of the freely suspended film of thickness h_0 .

The laser beam in the interferometer passes twice through the film, as indicated, to measure the electrically induced thickness modulation Δh . The electric field E is applied across the whole, 1-mm-wide film, perpendicular to the laser beam and Δh directions. **c**, The electroclinic effect. The viewing angle is turned by 90° around the layer normal compared to that in **b**. Without electric field, one smectic layer has a thickness h_0/m , if the whole film consists of m layers, and the mesogenic side groups stand (on average) upright ($\theta = 0^\circ$) inside the single smectic layers (S_A^* phase). By application of a lateral electric field, a tilt angle θ , proportional to the electric field E , is induced in a plane perpendicular to the field. The sign of the tilt depends on whether the electric field E points out of (circled dot) or into (circled cross) the plane of the page. Hence, each smectic layer shrinks by $\Delta h/m$ two times during one period of the a.c. electric field.

they just flow. This problem can be overcome by the synthesis of liquid-crystalline polymers, where the mesogens are attached, in a comb-like fashion, to a polysiloxane backbone via flexible alkyl spacers (Fig. 1a). The resulting side-chain polymer may then be crosslinked, which yields a liquid-crystalline elastomer (LCE, rubber)⁵. Moreover, in these liquid-crystalline polymers, the mesogenic side chains can still be arranged in a self-organized, layered ('smectic') structure (Fig. 1b). If crosslinked then, a true one-dimensional, long-range order can be established and preserved on a macroscopic length scale⁶ in the resulting LCE. The soft

elastomeric network stabilizes the order that was present in the self-organized smectic films of the polymer during the crosslinking, but leaves at the same time sufficient mobility for electrically induced molecular reorientations of the mesogenic side chains against the restoring elastic forces of the network.

The idea of the present approach to achieve high electromechanical responses in a smart material is to make use of ferroelectric LCEs (FLCEs). In contrast to other LCEs, FLCEs are built mainly from chiral mesogens. Details of FLCEs may be found in refs 7–10; here we can only give a brief description of where the ferroelectricity in ferroelectric liquid crystals (FLCs) originates. It is a consequence of the special symmetry of the tilted, ferroelectric S_C^* phase in these samples (Fig. 2b, top): consider the rotation of a non-chiral, tilted mesogen by π (C_2 -symmetry operation) around one of its short axes—the one (coinciding with the $\hat{y}$ -axis in Fig. 1b) that is perpendicular to the tilt plane (the $\hat{x}$ – $\hat{z}$ plane in Fig. 1c). Then we apply a mirror operation (σ_{xz}) with the tilt plane as mirror plane. Both C_2 and σ_{xz} are allowed, because they do not alter the image in Fig. 2b. Nevertheless, as a result, every permanent dipole moment $\mathbf{P}$ of a mesogen can be turned into its negative by consecutive application of the two symmetry operations: $\mathbf{P} = (p_x, p_y, p_z) \xrightarrow{C_2} (-p_x, p_y, -p_z) \xrightarrow{\sigma_{xz}} (-p_x, -p_y, -p_z) = -\mathbf{P}$. Hence, the net polarization of a number of non-chiral mesogens cancels. If, on the other hand, the liquid crystal consists of chiral—this means only right-handed (or only left-handed)—versions of the mesogens, as in FLCs, the mirror operation σ_{xz} is no longer a symmetry operation of this phase. Therefore, a net polarization may remain in the y direction, perpendicular to the ($\hat{x}$ – $\hat{z}$) tilt plane, if additionally the thermally induced rotation of the mesogens around their long axes is biased anisotropically owing to the shape of the molecules.

In contrast, in the paraelectric S_A^* phase (Fig. 2c, top) of FLCs the long axes of the mesogens are aligned parallel to the layer normals (tilt angle $\theta = 0^\circ$). However, if the molecular tilt is regarded as the cause of the spontaneous polarization in the ferroelectric S_C^* phase, an externally induced electric polarization may in turn be used to induce a tilt θ of the mesogenic side chains in the paraelectric S_A^* phase. This electrically induced tilt ('the electroclinic effect') of the mesogens is proportional to the externally applied electric field¹¹ (Fig. 1c). It leads to a constriction of the smectic layers, as measured already by X-ray diffraction on a low-molecular-weight (M_w) FLC¹². The layer constriction may be understood from geometrical considerations, because the layer spacing is proportional to the cosine of the tilt angle θ . A Taylor expansion of the cosine around $\theta = 0^\circ$ leads to a proportionality of the constriction of one single layer $\Delta h/m$ to the square of the electric field E , which indicates electrostriction. Consequently, the layer constriction occurs at twice the frequency, the 'second harmonic', of an externally applied alternating current (a.c.) electric field E_{ac} .

In our experiments we measured the thickness variation Δh of ultrathin (< 100 nm) substrate-free FLCE films (Fig. 2a) in response to a.c. electric fields. The films were crosslinked in the S_A^* phase by irradiation with ultraviolet light¹³, which yields a truly solid elastic (elastic modulus ~ 3 MPa, see Supplementary Information)

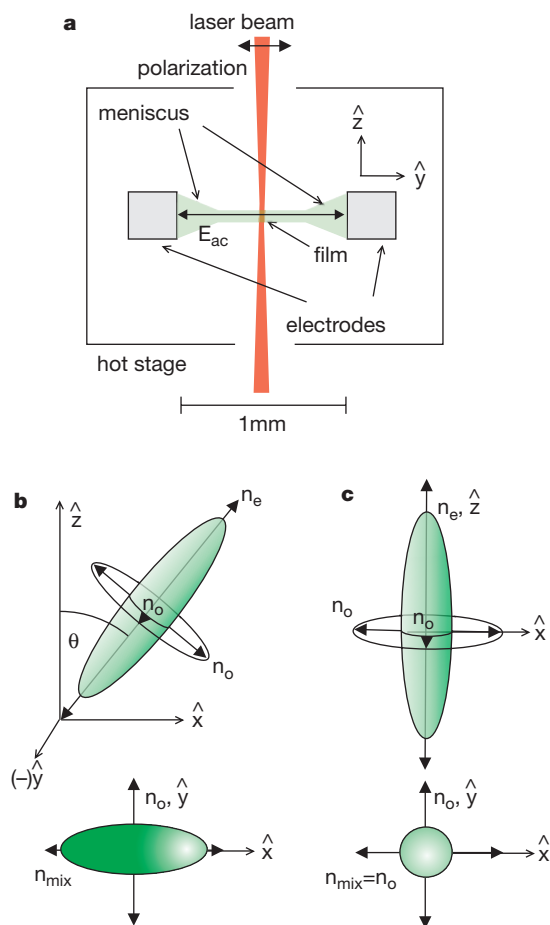


Figure 2 Optical geometries. **a**, Sketch of the measurement geometry. The coordinate system shown applies to all panels in this figure. The thickness of the film is exaggerated. The smectic layers are parallel to the plane of the film (= homeotropic orientation). The film spans the distance between two electrodes (the applied electric field is denoted as E_{ac}), right in the common focal plane of two long-distance microscope lens systems above and below the film, far out of the frame of this picture. This allows for a small measuring spot (100 μm) and a negligible aperture angle. **b**, Optical axes in the tilted (biaxial) state. Top panel shows side view of the rod-like liquid-crystal molecule(s), showing the orientation of the optical axes. The ordinary refractive index n_o applies to all directions perpendicular to the molecules' long axes, and the extraordinary refractive index n_e applies parallel to the molecules' long axes. Bottom panel shows top view. To a very good approximation, the biaxial film can be treated as a uniaxial film having the indices n_o and n_{mix} . The latter is a mixture of both n_o and n_e , and is a function of the tilt angle θ . It can be calculated via the ellipsoidal relation $1/n_{mix}^2 = (\cos^2\theta/n_o^2 + \sin^2\theta/n_e^2)^{1/2}$. **c**, Optical axes in the non-tilted (uniaxial) state. Top panel shows side view. Bottom panel shows top view; the viewing direction is the optical axis of the film. The mixed index n_{mix} is now identical to n_o , which can also be found on setting $\theta = 0^\circ$ in the ellipsoidal relation above. It is evident from the bottom parts of **b** and **c** that if the polarization of the laser beam is chosen to be parallel to the $\hat{y}$ direction (which is also the direction of the applied external electrical field E_{ac}), the relevant refractive index is always n_o and is independent of the electrically induced tilt angle θ .

Table 1 The achievable strains in various electrostrictor materials

Material	Electrostrictive strain* ϵ (at electric field strength E)	Ref.
Freely suspended FLCE film	4% lateral strain (1.5 MV m ⁻¹)	This work
P(VDF-TrFE): electron-irradiated	4%	4
poly(vinylidene fluoride-trifluoroethylene) copolymer	(150 MV m ⁻¹)	
PMN _{0.7} PT _{0.3} : lead magnesium niobate-lead titanate ceramics	0.15% (1.0 MV m ⁻¹)	17
PBLG: monomolecular, grafted layer of poly- γ -L-glutamate	0.005% (300 MV m ⁻¹)	1

* $\epsilon = \Delta h/h_0$.

membrane¹⁴ in contrast to the free-standing films of liquid-like low- M_w liquid crystals—for example, those of refs 15, 16. For the measurement of the electrically induced change Δh in sample thickness, we used a high-precision (± 3 pm at the frequency $\omega = 133$ Hz of the applied electric field) Michelson interferometer in transmission mode, which was designed especially for this purpose and allows us to distinguish the desired optical phase shift from all the other (time-dependent) optical effects in the birefringent FLCE (see Methods). In substrate-free FLCE films the smectic layers are arranged parallel to the plane of the film, which means that the constrictions $\Delta h/m$ of all m layers sum up to the overall constriction Δh of the film. The electric field is applied to the lateral electrodes with a gap of 1 mm, between which the film is freely suspended (Fig. 2a). Hence, the FLCE film builds a self-assembled actuator stack needing only two lateral electrodes.

The measured data depicted in the main panel of Fig. 3 show that the linear (piezoelectric) contribution at the first harmonic (ω) is small. It can be attributed to tiny angular mismatches in the set-up—such as may be caused if the laser beam is not perfectly perpendicular to the film plane—or to a residual ferroelectric contribution. In contrast, the observed electrostriction at the second harmonic (2ω), which follows well the expected square dependence on U_{ac} , is huge: the FLCE film is thinning by 4% at a lateral electric field E_{ac} (that is, U_{ac} divided by the 1-mm width of the electrode gap) of only 1.5 kV mm^{-1} , which yields the largest value of the electrostriction coefficient¹ a (defined by $\Delta h/h_0 = (1/2)aE_{ac}^2$) yet reported. Table 1 gives an overview of the highest achievable strains and the required electric field strengths in FLCE and in competing electrostrictor materials^{1,4,17}. The electric field strength needed for a strain of 4% is 100 times lower in FLCE than in the second-best electrostrictive material, poly(vinylidene fluoride-trifluoroethylene) (ref. 4). The film thinning of 4% in the FLCE corresponds to an electrically induced tilt angle $\theta \approx 16^\circ$ in the simple cosine

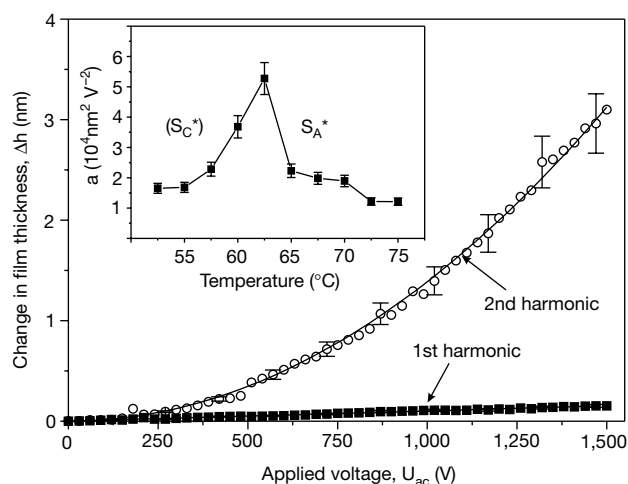


Figure 3 Electromechanical responses in FLCEs. Main panel, electrically induced thickness variations Δh of a freely suspended FLCE film as a function of the applied lateral a.c. voltage with amplitude U_{ac} ($\omega = 133$ Hz) at the first harmonic (filled squares, error bars smaller than the symbols) and at the second harmonic (open circles), as determined in the interferometer at $T = 60^\circ\text{C}$. The film thickness h_0 of 75 ± 5 nm was determined by reflectivity measurements. Film width equals the inter-electrode distance of 1 mm. Inset, electrostrictive coefficient a (filled squares) versus temperature T . The coefficient a was derived from the square dependence of Δh on the electric field, as noted in the text. At the Curie temperature of the non-crosslinked polymer (phase transition $S_C^* \rightarrow S_A^*$, $T_C \approx 61.0^\circ\text{C}$), the tilt angle θ is a soft parameter. Therefore, the electroclinic effect is most pronounced at this temperature. In the figure, the symbol S_C^* is in brackets because the crosslinking stabilizes ('supercools') the S_A^* -like ordering even below the $S_C^* \rightarrow S_A^*$ phase-transition temperature of the uncrosslinked system.

model. This is a reasonable number, as comparable θ values have already been measured optically for an FLC-polysiloxane in a commercial, 4- μm -thick sample cell¹⁸. The inset in Fig. 3 shows the electrostriction coefficient a as derived from fits to the second-harmonic data at different temperatures. The electrostrictive response has a pronounced maximum near the phase transition of the uncrosslinked FLCP from ferroelectric (S_C^* phase) to paraelectric (S_A^* phase), at the Curie temperature $T_C \approx 61^\circ\text{C}$. This proves the correlation with the electroclinic effect: θ is a soft parameter at the phase transition, which means that a tilt can be most easily induced by an electric field at that temperature where a spontaneous breaking of symmetry occurs upon cooling. The electroclinic origin of the effect was also proved by X-ray diffraction measurements (data not shown) on FLCE films that were spin-cast on a glass substrate; these measurements showed clearly the electroclinic layer constriction.

This flexible, transparent and highly electrostrictive material is well suited for use in nanomachines¹⁹, for medical applications, for example. FLCEs are also useful in other applications. We reported recently that the piezoelectric moduli of different types of FLCEs²⁰ can be higher than those of commercial piezoelectrics, such as lead zirconate titanate, or poled polymers such as poly(vinylidene fluoride)^{21–23}. Nevertheless, we are aware of no piezoelectric material that can produce strains of the order of magnitude of those reported here for the electrostrictive FLCE. □

Methods

We used a Michelson interferometer similar to that in ref. 21 to measure the periodic, electrically induced thickness variation Δh of the film as a shift Ξ (in nm) in optical path length between the sample beam, passing the sample twice (Fig. 1b, Fig. 2a), and a reference beam. The sample is mounted at the common focal plane of two long-distance microscope lens systems that are placed above and below the film. This allows for a small measurement spot ($\sim 100 \mu\text{m}$) and a negligible aperture angle. The mutual orientation of the electric field and the laser polarization is chosen to be parallel. In this geometry the laser beam is affected by the ordinary refractive index n_0 (≈ 1.51) of the film only, which means that the relevant refractive index is not a function of θ (see Fig. 2). The considerations above apply if conservation of sample density is assumed. This means that the film has to expand in the $\hat{x}$ and $\hat{y}$ directions, when it shrinks in the $\hat{z}$ direction. We know that this is the case, because if no material moved sideways out of the laser spot, there would have been nothing to measure: the increase in sample density (and thus in refractive index) would compensate for the reduced thickness of the film and the optical path length would not change. Hence, the change in film thickness Δh can be calculated from the measured difference in optical path length Ξ as follows: $\Xi = h_0 n_0 - (\Delta h n_0 + (h_0 - \Delta h) n_0) = \Delta h(n_0 - 1)$. Nevertheless, it is still necessary to exclude the possibility that scattering, reflections or a modulation of the state of polarization of the laser light contribute to the measured signal. To do so, we put an analyser in the recombined beam, in front of the detector. All the measurements were then performed twice: during the second measurement the reference beam of the interferometer was blocked out and the undesired (electrooptical) modulations were measured separately. They were found to be negligible.

Received 12 July; accepted 13 December 2000.

- Jaworek, T., Neher, D., Wegner, G., Wieringa, R. H. & Schouten, A. J. Electromechanical properties of an ultrathin layer of directionally aligned helical polypeptides. *Science* **279**, 57–60 (1998).
- Pelrine, R., Kornbluh, R., Pei, Q. & Joseph, J. High-speed electrically actuated elastomers with strain greater than 100%. *Science* **287**, 836–839 (2000).
- Baughman, R. H. et al. Carbon nanotube actuators. *Science* **284**, 1340–1344 (1999).
- Zhang, Q. M., Bharti, V. & Zhao, X. Giant electrostriction and relaxor ferroelectric behavior in electron-irradiated poly(vinylidene fluoride-trifluoroethylene) copolymer. *Science* **280**, 2101–2104 (1998).
- Zentel, R. Liquid crystalline elastomers. *Angew. Chem. Int. Edn Engl.* **28**, 1407–1415 (1989).
- Wong, G. C. L. et al. Induced long range order in crosslinked 'one-dimensional' stacks of fluid monolayers. *Nature* **389**, 576–579 (1997).
- Lagerwall, S. T. *Ferroelectric and Antiferroelectric Liquid Crystals* (Wiley-VCH, Weinheim, 1999).
- Terentjev, E. M. & Warner, M. Continuum theory of ferroelectric smectic C^* elastomers. *J. Phys. II France* **4**, 849–858 (1994).
- Clark, N. A. & Lagerwall, S. T. Submicrosecond bistable electro-optic switching in liquid crystals. *Appl. Phys. Lett.* **36**, 899–901 (1980).
- Jákli, A. & Saupe, A. Field-induced thickness change of ferroelectric liquid crystal films. *Phys. Rev. E* **53**, R5580–R5583 (1996).
- Garoff, S. & Meyer, R. B. Electroclinic effect at the A-C phase change in a chiral liquid crystal*. *Phys. Rev. Lett.* **38**, 848–851 (1977).
- Rappaport, A. G. et al. X-ray observation of electroclinic layer constriction and rearrangement in a chiral smectic-A liquid crystal. *Appl. Phys. Lett.* **67**, 362–364 (1995).
- Gebhard, E. & Zentel, R. Ferroelectric liquid crystalline elastomers. *Macromol. Chem. Phys.* **201**, 902–910 (2000).

14. Schüring, H., Stannarius, R., Tolkdorf, C. & Zentel, R. Liquid crystal elastomer balloons. *Macromolecules* (submitted).
15. Mach, P., Huang, C. C. & Nguyen, H. T. Dramatic effect of single-atom replacement on the surface tension of liquid-crystal compounds. *Phys. Rev. Lett.* **80**, 732–735 (1998).
16. Johnson, P. M., Pankratz, S., Mach, P., Nguyen, H. T. & Huang, C. C. Optical reflectivity and ellipsometry studies of the Sm-C_a* phase. *Phys. Rev. Lett.* **83**, 4073–4076 (1999).
17. Zhao, J., Zhang, Q. M., Kim, N. & Shrout, T. Electromechanical properties of relaxor ferroelectric lead magnesium niobate-lead magnesium titanate ceramics. *Jpn J. Appl. Phys.* **34**, 5658–5663 (1995).
18. Poths, H., Anderson, G., Skarp, K. & Zentel, R. Fast electroclinic switching in a ferroelectric LC Polysiloxane. *Adv. Mater.* **4**, 792–794 (1992).
19. Kremer, F. Electromechanical and/or mechanoelectrical converter. German Patent No. 196 36 909 (1998).
20. Semmler, K. & Finkelmann, H. Mechanical field orientation of chiral smectic C-polymer networks. *Macromol. Chem. Phys.* **196**, 3197–3205 (1995).
21. Lehmann, W. *et al.* The electromechanical effect in mechanically oriented S_C*-elastomers examined by means of an ultra-stable Michelson interferometer. *Ferroelectrics* **208–209**, 373–383 (1998).
22. Lehmann, W. *et al.* Direct and inverse electromechanical effect in ferroelectric liquid crystalline elastomers. *J. Appl. Phys.* **86**, 1647–1652 (1999).
23. Kremer, F. *et al.* Piezoelectricity in ferroelectric liquid crystalline elastomers. *Polym. Adv. Technol.* **9**, 672–677 (1998).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank R. Stannarius and D. Neher for discussions, and the "Innovationskolleg Phänomene an den Miniaturisierungsgrenzen" at the University of Leipzig for support.

Correspondence and requests for materials should be addressed to F.K. (e-mail: kremer@physik.uni-leipzig.de).

Evolution of nanoporosity in dealloying

Jonah Erlebacher^{*†}, Michael J. Aziz^{*}, Alain Karma[‡], Nikolay Dimitrov[§] & Karl Sieradzki[§]

^{*} Division of Engineering and Applied Sciences, Harvard University, 9 Oxford Street, Cambridge, Massachusetts 02138, USA

[‡] Department of Physics and Center for Interdisciplinary Research on Complex Systems, Northeastern University, 360 Huntington Avenue, Boston, Massachusetts 02115, USA

[§] Department of Mechanical and Aerospace Engineering and Center for Solid State Sciences, Arizona State University, Tempe, Arizona 85287-6106, USA

Dealloying is a common corrosion process during which an alloy is 'parted' by the selective dissolution of the most electrochemically active of its elements. This process results in the formation of a nanoporous sponge composed almost entirely of the more noble alloy constituents¹. Although considerable attention has been devoted to the morphological aspects of the dealloying process, its underlying physical mechanism has remained unclear². Here we propose a continuum model that is fully consistent with experiments and theoretical simulations of alloy dissolution, and demonstrate that nanoporosity in metals is due to an intrinsic dynamical pattern formation process. That is, pores form because the more noble atoms are chemically driven to aggregate into two-dimensional clusters by a phase separation process (spinodal decomposition) at the solid–electrolyte interface, and the surface area continuously increases owing to etching. Together, these processes evolve porosity with a characteristic length scale predicted by our continuum model. We expect that chemically tailored nanoporous gold made by dealloying Ag–Au

should be suitable for sensor applications, particularly in a biomaterials context.

Selective dissolution has a long history³. For example, the chemical treatment known as depletion gilding selectively dissolves a non-gold element near the surface of a less-expensive alloy such as Au–Cu, leaving behind a surface of pure gold. Early Andean goldsmiths used this technique to enhance the surfaces of their artefacts⁴. In this century, selective dissolution has been primarily examined in the context of corrosion. It is observed in technologically important alloy systems, notably brasses, stainless steels, and Cu–Al alloys^{1,5,6}. The mechanical properties of a porous overlayer are very different from the bulk alloy on which it sits, leading to brittle crack propagation, stress corrosion cracking, and other undesirable materials failure⁷. Figure 1 shows the prototypical dealloyed microstructure, that of nanoporous gold. Early notions considered porosity as a hidden microstructure revealed by etching, but diffraction experiments showed that no pre-existing length scale exists before acid attack on single-phase alloys^{8,9}. Later ideas considered the influence of percolating clusters within the solid solution of the alloy, but models failed to yield behaviour consistent with experiment^{10,11}.

The following argument illustrates the fundamental obstacle to understanding porosity formation during dealloying: consider a silver–gold alloy in an electrolyte under conditions where silver dissolves and gold is inert. Initially, silver will be dissolved from surface sites such as terraces or steps. Gold atoms should accumulate on the surface and locally block further dissolution. For an alloy containing 10% gold, it might be expected that dissolution would stop or be significantly retarded after about 10 monolayers of the alloy have been dissolved.

A complete model of selective dissolution needs to be multi-scale, involving the kinetics of dissolution, surface diffusion, and mass transport through the bulk of both alloy and electrolyte. Because

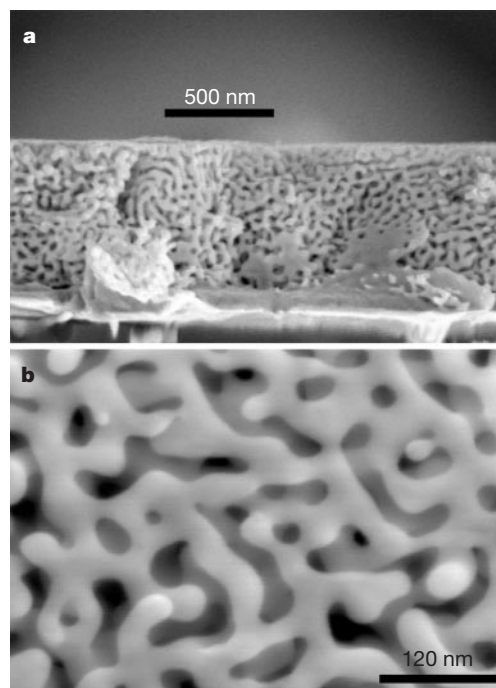


Figure 1 Scanning electron micrographs of nanoporous gold made by selective dissolution of silver from Ag–Au alloys immersed in nitric acid under free corrosion conditions. **a**, Cross-section of dealloyed Au₃₂Ag₆₈ (atom%) thin film. **b**, Plan view of dealloyed Au₂₆Ag₇₄ (atom%). The porosity is open, and the ligament spacings shown in **b** are of the order of 10 nm; spacings as small as 5 nm have been observed. Measurements of the surface area of nanoporous gold are of the order of order 2 m² g^{−1} (refs 24, 25), comparable to commercial supported catalysts.

[†] Present address: Department of Materials Science and Engineering, Johns Hopkins University, Baltimore, Maryland 21218, USA.

mass transport through the bulk of the growing phase (the electrolyte) is always a stabilizing influence¹², and because mass transport through the bulk of the dissolving phase appears too slow to be significant, we hypothesized that the morphology-determining physical process is confined to the interface region between the alloy and the electrolyte. To test this, we developed a kinetic Monte Carlo model to simulate Ag-Au dealloying, including only diffusion of silver and gold and dissolution of silver¹³. We found that this model was able to reproduce all relevant experimental trends characteristic of dealloying, both morphological and kinetic.

Figure 2 shows a simulated porous structure with ligament widths of 2–5 nm. Our simulations were successful in modelling the nanoporous morphology, and also in modelling the dynamic behaviour of the dissolution current versus overpotential. It is a well characterized feature of alloy dissolution that as the overpotential is increased (usually at rates of the order of a few millivolts per second), the dissolution current of ions from the alloy stays at a low level until a bulk-composition-dependent critical potential (V_C) is reached, at which point this current rises rapidly¹⁴. Figure 3 shows simulated and experimental polarization curves for different alloy compositions. There is clear observation of a composition-dependent V_C . To our knowledge, this is the first simulation model to produce such behaviour, suggesting that we have found a minimum set of physical processes to include in any model for

alloy dissolution.

The simulations reveal the following qualitative picture of porosity formation. The process starts with the dissolution of a single silver atom on a flat alloy surface of close-packed (111) orientation, leaving behind a terrace vacancy. The atoms coordinating this vacancy have fewer lateral near-neighbours than other silver atoms in the terrace, and are thus more susceptible to dissolution. As a result, the entire terrace is 'stripped', leaving behind gold atoms with no lateral coordination ('adatoms'). Before the next layer is attacked, these gold adatoms diffuse about and start to agglomerate into islands. As a result, rather than a uniform diffuse layer of gold spread over the surface, the surface is comprised of two distinct kinds of regions—pure gold clusters that locally passivate the surface, and patches of un-dealloyed material exposed to electrolyte. When silver atoms in these patches dissolve, more gold adatoms are released onto the surface. These adatoms diffuse to the gold clusters left over from dissolution of previous layers, continuing to leave un-dealloyed material exposed to electrolyte. In the early stages, these gold clusters are mounds that are gold-rich at their peaks and that have alloy composition at their bases. These mounds get undercut, increasing the surface area that gold must cover to bring about passivation. Ultimately, this leads to pit formation and porosity.

Central to this description is the the coalescence of gold adatoms into stable clusters. The spacing between these 'islands' in the initial stages of dissolution is close to the spacing between ligaments in the final porous structure. The physical reason for this coalescence can be understood by considering the gold adatoms to be one component of a two-component solution of gold and 'electrolyte' confined to the monolayer-thick interfacial layer sitting on top of un-dealloyed material. We modelled the thermodynamics of the interfacial layer as a regular solution¹⁵, and found the solubility of gold in electrolyte within the interfacial layer to be of the order of 10^{-7} per site (see Methods). This solubility may be interpreted as the 'equilibrium concentration of gold adatoms' on the surface of the alloy—in the absence of etching, it represents a dynamic equilibrium of adatoms resulting from their two-dimensional evaporation from step edges onto terraces and their subsequent recondensation.

In contrast to the equilibrium condition, rapidly stripping a terrace of silver atoms leaves gold adatoms with a local site occupancy fraction equal to that in the bulk, typically 10–40%—far above their equilibrated concentration of 10^{-7} per site. Thus, there is an extremely strong driving force for gold adatoms to condense onto nearby gold-rich clusters. In fact, regions of surface with high enough supersaturation of gold adatoms sit 'within the spinodal', a special segment of the curve representing free energy f of a spatially uniform layer versus gold concentration C for which $\partial^2 f / \partial C^2 < 0$. Within the spinodal, composition fluctuations of infinitesimal amplitude lead to a lower overall free energy for the

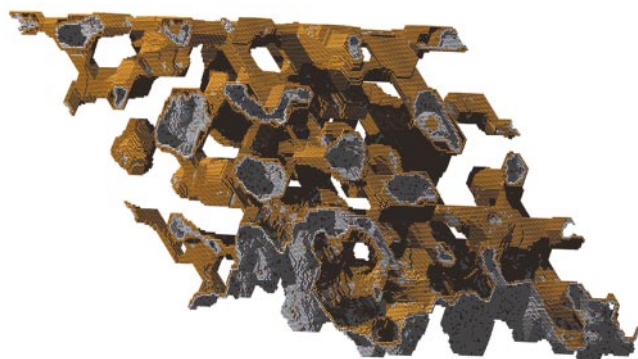


Figure 2 Simulated nanoporous gold. The simulation model was as follows: a bond-breaking model was used for diffusion; atoms with N near neighbours diffused with rate $k_N = \nu_D \exp(-N\epsilon/k_B T)$, where ϵ is a bond energy and $\nu_D = 10^{13} \text{ s}^{-1}$. Dissolution rates were consistent with the Butler–Volmer (BV) equation in the high-driving-force Tafel regime; the dissolution rate $k_{E,N}$ for a silver atom with N near neighbours was written as $k_{E,N} = \nu_E \exp[-(N\epsilon - \phi)/k_B T]$, where $\nu_E = 10^4 \text{ s}^{-1}$ is an attempt frequency determined by the exchange-current density in the BV equation and ϕ is the overpotential. For the figure, $\phi = 1.75 \text{ eV}$, $\epsilon/k_B T = 5.51$.

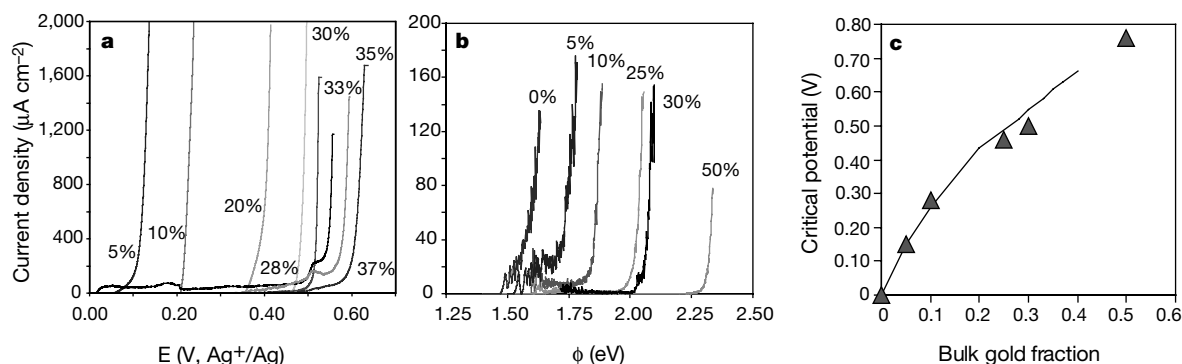


Figure 3 Comparison of experimental and simulated current–potential behaviour. **a**, Current–potential behaviour for varying Ag-Au alloy compositions (atom% Au) dealloyed in 0.1 M HClO_4 + 0.1 M Ag^+ (reference electrode 0.1 M Ag^+/Ag). **b**, Simulated

current–potential behaviour of Ag-Au alloys. **c**, Comparison of experimental (line) and simulated (triangles) critical potentials; the zero of overpotential has been set equal to the onset of dissolution of pure silver both in simulation and in experiment.

system, and involve atomic diffusion against concentration gradients (the 'uphill diffusion' process through which gold condenses onto nearby clusters), that is, the system is inherently unstable and will spontaneously phase-separate. But fluctuations of long length scale grow slowly due to the required diffusion times, and short-length-scale fluctuations create much energetically unfavourable incipient interface between the phases, inhibiting their growth. Hence, phase separation is manifested most rapidly at an intermediate length scale that roughly corresponds to the spacing between the observed gold-rich clusters. This effect is known as spinodal decomposition^{16,17}. As porosity forms, the decomposition is occurring on a non-flat, non-uniform surface with continuously increasing surface area.

In our model, the motion of the alloy–electrolyte interface is fully described mathematically by the flux of diffusing adatoms J_s , the velocity of the interface normal to itself v_n , and the concentration accumulation rate $\partial C/\partial t$, all of which are interrelated and vary with position along the curve of the interface (for detailed derivations, see Methods). For J_s , we used a model for diffusion during spinodal decomposition known as the Cahn–Hilliard equation¹⁷. The normal velocity depends on C and also on the local curvature κ through capillary effects¹¹. The time evolution of C is uniquely determined by the local mass-conservation condition

$$\partial C/\partial t = v_n C_0 - v_n \kappa C - \nabla \cdot J_s \quad (1)$$

where C_0 is the bulk gold concentration. This condition is analogous to the local conservation of heat or solute appearing in boundary-layer models of solidification¹⁸, with two important differences: (1) the interfacial layer thickness is constant along the interface, and is microscopic, rather than being a spatially varying macroscopic diffusion length, and (2) the surface aggregation process inherent in the Cahn–Hilliard form for J_s is essential for porosity formation. Simple ('downhill') surface diffusion ($J_s = -D_s \nabla C$) yields an

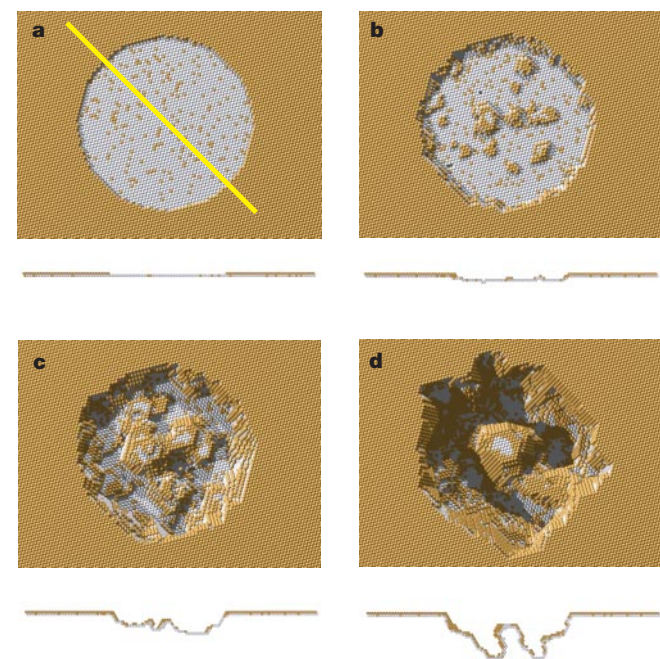


Figure 4 Simulated evolution of an artificial pit in Au_{10%}Ag_{90%} (atom%), $\phi = 1.8$ eV. Cross-sections along the (111) plane defined by the yellow line in **a** are shown below each plan view. **a**, The initial condition is a surface fully passivated with gold except within a circular region (the 'artificial pit'). **b**, After 1 s, the pit has penetrated a few monolayers into the bulk. We note how there are fewer gold clusters near the side wall than at the centre of the pit. **c**, After 10 s, a gold cluster has nucleated in the centre of the pit. **d**, At 100 s, the pit has split into multiple pits; each will continue to propagate into the bulk to form a porous structure like that in Fig. 2.

initially unstable interface that passivates quickly, before well-formed pores have a chance to develop. Cahn–Hilliard diffusion also dominates capillarity-driven surface diffusion—the effect usually incorporated into interface evolution equations¹⁹.

We performed numerical integration of equation (1) using a relative arc-length parametrization scheme^{20,21}, and parameters that matched those used in the kinetic Monte Carlo simulations. We observed, as expected, the evolution of gold clusters separated by a characteristic spacing λ . An analytic expression for λ can be found by a time-dependent linear stability analysis of equation (1) that takes into account the slow increase of gold concentration into the interfacial layer as the instability develops. This effect needs to be included because the spatial period with the largest amplification rate depends sensitively on the gold concentration. Specifically, this spatial period decreases sharply as the concentration increases past a threshold concentration for instability that corresponds to the spinodal point $\partial^2 f/\partial C^2 = 0$, and the interface is stable for concentrations below this threshold. We find a maximally unstable spatial period that scales as $\lambda \propto (D_s/V_0)^{1/6}$, where V_0 is the velocity of a flat alloy surface with no gold accumulated upon it. This prediction is in qualitative agreement with both kinetic Monte Carlo simulations and experiments, both of which show that the characteristic length scale of porosity decreases with increasing driving force. A more elaborate analysis incorporating nonlinear effects, however, remains needed for a detailed quantitative comparison.

There is an analogy between this result, applicable to etching, and two-dimensional island nucleation during submonolayer vapour phase deposition. Namely, in the early stages of etching, the dissolution process is analogous to deposition of gold; in both processes, adatoms are added to the surface where they are free to agglomerate into islands. The case of vapour phase deposition has been studied using rate equations that describe an aggregation process where adatoms stick together irreversibly. In these studies it is a well-known result that the island spacing scales as $(D_s/F)^\mu$, where the deposition rate F is the direct analogue of the surface velocity in etching, and the exponent μ depends on details of the aggregation process (see ref. 22 and references therein). That these results are limited to irreversible aggregation during deposition and our analysis is for reversible aggregation during etching suggests the existence of universal scaling laws for aggregation that do not depend on reversibility or the lack thereof in these two opposite processes.

Our kinetic Monte Carlo simulation elucidates the later stages of morphological evolution, and the mechanism by which three-dimensional porosity evolves. We highlight the features of this process by showing in Fig. 4 a simulation of an artificial pit in an otherwise fully passivated surface. When the pit reaches sufficient depth, its surface area has increased sufficiently that a new gold cluster nucleates. When this happens, the pore splits into multiple new pits, each with a smaller surface area than its parent. These 'child' pits continue to penetrate into the bulk, increasing their surface area, nucleating new clusters, spawning new pits, and so on, until a full, three-dimensional porous structure evolves, such as those illustrated in Figs 1 and 2. □

Methods

In a regular solution, the enthalpy of mixing depends on the bond energies and the entropy of mixing is ideal. The free energy of a regular solution $f(C, T)$ is written $f = \alpha c(1 - c) + k_B T [c \ln c + (1 - c) \ln(1 - c)]$, where c is the mole fraction of gold ($c = C\Omega^3$, where Ω is atomic volume), $\alpha = 6[E_{\text{Au-electrolyte}} - (1/2)(E_{\text{Au-Au}} + E_{\text{electrolyte-electrolyte}})]$, where E_{x-x} are the respective interaction energies between Au and electrolyte, the prefactor 6 is the lateral coordination in the two-dimensional hexagonal lattice of the interfacial layer, k_B is Boltzmann's constant and T is absolute temperature. For our simulation conditions, realistic timescales and length scales were obtained from the parameters $E_{\text{Au-Au}} = -0.285$ eV ($= -\epsilon$, the simulation bond energy as described in Fig. 2), $T = 600$ K, $E_{\text{Au-electrolyte}} = E_{\text{electrolyte-electrolyte}} = 0.0$ eV. With these parameters, $\alpha = 0.855$ eV. The free energy has the familiar double-well form²³ and a minimum at $c \approx 10^{-7}$ per site, representing the solubility of gold in electrolyte (and vice versa).

The Cahn–Hilliard diffusion equation is $J_s = -M(C)(\partial^2 f/\partial C^2) \nabla C + 2M(C)w \nabla^3 C$.

Here, $453M(C)$ is a mobility, w is the so-called gradient energy coefficient, and the gradients are taken with respect to arc length. The first term on the right-hand side describes the chemical effect leading to phase separation within the spinodal; the second term describes the effect that damps short-wavelength fluctuations. The mobility is proportional to the surface diffusivity D_s and is given by $M(C) = (D_s/k_B T)c(1-c)$. The mobility is peaked for $c = 0.5$, and is zero for $c = 0$ and $c = 1$ (atoms do not diffuse in pure phases because there are no vacancies in our model). The normal velocity is given by $v_n(C) = V(C)[1 - (\gamma\Omega/k_B T)\kappa]$, where γ is the surface free energy and $V(C)$ is called the interface response function, equal to the velocity of a flat surface covered with a concentration C of gold. We find in both simulation and experiment that the interface response is fitted well by the functional form $V(C) = V_0(\phi)\exp(-C/C^*)$, where ϕ is the overpotential and C^* is a constant. Experimentally, the gold accumulation can be inferred by integrating the dissolution current versus time at fixed overpotential; it is necessary to use an overpotential that is low enough to ensure that the surface remains planar (that is, porosity does not form) and also to catch the short initial transient rise in current as silver atoms are pulled from the first few monolayers. This particular form for the interface response function is quite curious. Naively, one might expect that the local interface velocity would be proportional to the local concentration of silver exposed to the electrolyte, that is, $V(C) \propto (1-c)$. However, the decaying exponential form suggests that there is an evolving distribution of holes opening and closing within the interfacial region, controlling the accumulation rate.

Physically, the mass conservation condition (equation (1)) is the statement that the total number $Cb\Delta s$ of gold atoms in a length Δs of interface with lateral width b can change as a result of three distinct effects that correspond to the three terms on the right-hand-side of equation (1): the accumulation of gold atoms into the interfacial layer from the solid being dissolved; the local stretching of the interface ($\partial\Delta s/\partial t = v_n\kappa\Delta s$), which can either increase or decrease C depending on whether the solid is concave ($\kappa > 0$) or convex ($\kappa < 0$); and the motion of atoms along the interface driven by the surface diffusion flux J_s .

Received 14 November 2000; accepted 10 January 2001.

- Pickering, H. W. Characteristic features of alloy polarization curves. *Corros. Sci.* **23**, 1107–1120 (1983).
- Forty, A. J. Corrosion micromorphology of noble metal alloys and depletion gilding. *Nature* **282**, 597–598 (1979).
- Masing, G. Zur Theorie der Resistenzgrenzen in Mischkristallen. *Z. Anorg. Allg. Chem.* **118**, 293–308 (1921).
- Lechtman, H. Pre-Columbian surface metallurgy. *Sci. Am.* **250**, 56–63 (1984).
- Williams, D. E., Newman, R. C., Song, Q. & Kelly, R. G. Passivity breakdown and pitting corrosion of binary alloys. *Nature* **350**, 216–219 (1991).
- Newman, R. C. & Sieradzki, K. Metallic Corrosion. *Science* **263**, 1708–1709 (1994).
- Li, R. & Sieradzki, K. Ductile-brittle transition in random porous Au. *Phys. Rev. Lett.* **68**, 1168–1171 (1992).
- Corcoran, S. in *Critical Factors in Localized Corrosion III* (eds Kelly, R. G., Frankel, G. S., Natishan, P. M. & Newman, R. C.) 500–507 (Electrochemical Society, Pennington, New Jersey, 2000).
- Pickering, H. W. & Wagner, C. Electrolytic dissolution of binary alloys containing a noble metal. *J. Electrochem. Soc.* **114**, 698–706 (1967).
- Sieradzki, K., Corderman, R. R., Shukla, K. & Newman, R. C. Computer simulations of corrosion: selective dissolution of binary alloys. *Phil. Mag.* **59**, 713–746 (1989).
- Sieradzki, K. Curvature effects in alloy dissolution. *J. Electrochem. Soc.* **140**, 2868–2872 (1993).
- Wagner, C. Contribution to the theory of electropolishing. *J. Electrochem. Soc.* **101**, 225–228 (1953).
- Erlebacher, J. in *Dynamics of Crystal Surfaces and Interfaces* (eds Duxbury, P. & Pence, T.) 24–28 (Plenum, New York, 1997).
- Wagner, K., Brankovic, S. R., Dmitrov, N. & Sieradzki, K. Dealloying below the critical potential. *J. Electrochem. Soc.* **144**, 3545–3555 (1997).
- Cahn, J. W. & Hilliard, J. E. Free energy of a nonuniform system. I. Interfacial free energy. *J. Chem. Phys.* **28**, 258–267 (1958).
- Cahn, J. W. & Hilliard, J. E. Free energy of a nonuniform system. III. Nucleation in a two-component incompressible fluid. *J. Chem. Phys.* **31**, 688–699 (1958).
- Hilliard, J. E. in *Solidification* 497–560 (American Society for Metals, Metals Park, Ohio, 1971).
- Ben-Jacob, E., Goldenfeld, N., Langer, J. S. & Schon, G. Dynamics of interfacial pattern formation. *Phys. Rev. Lett.* **51**, 1930–1932 (1983).
- Mullins, W. W. Theory of thermal grooving. *J. Appl. Phys.* **28**, 333–339 (1957).
- Kessler, D. A., Koplik, J. & Levine, H. Geometrical models of interface evolution. II. Numerical simulation. *Phys. Rev. A* **30**, 3161–3174 (1984).
- Hou, T. Y., Lowengrub, J. S. & Shelley, M. J. Removing the stiffness from interfacial flow with surface tension. *J. Comp. Phys.* **114**, 312–338 (1994).
- Barabási, A. -L. & Stanley, H. E. *Fractal Concepts in Surface Growth* (Cambridge Univ. Press, 1995).
- Haasen, P. *Physical Metallurgy* (Cambridge Univ. Press, 1986).
- Forty, A. J. in *Sir Charles Frank: An 80th Birthday Tribute* (ed. Chamber, R. G.) 164–187 (Adam Hilger, Bristol, 1991).
- Tulimieri, D. J., Yoon, J. & Chan, M. H. W. Ordering of helium mixtures in porous gold. *Phys. Rev. Lett.* **82**, 121–124 (1999).

Acknowledgements

This work was supported by the US Department of Energy, Basic Energy Sciences. The research of A.K. also benefited from computer time allocation at NU-ASCC. K.S. thanks the AFOSR for support.

Correspondence and requests for materials should be addressed to J.E (e-mail: Jonah.Erlebacher@jhu.edu).

Ice shelves in the Pleistocene Arctic Ocean inferred from glaciogenic deep-sea bedforms

Leonid Polyak*, Margo H. Edwards†, Bernard J. Coakley‡ & Martin Jakobsson§

* Byrd Polar Research Center, Ohio State University, Columbus, Ohio 43210, USA

† Hawaii Mapping Research Group, Hawaii Institute of Geophysics and Planetology, University of Hawaii, Honolulu, Hawaii 96822, USA

‡ Department of Geology, Tulane University, New Orleans, Louisiana 70118, USA

§ Department of Geology and Geochemistry, Stockholm University, 106 91 Stockholm, Sweden

It has been proposed that during Pleistocene glaciations, an ice cap of 1 kilometre or greater thickness covered the Arctic Ocean^{1–3}. This notion contrasts with the prevailing view that the Arctic Ocean was covered only by perennial sea ice with scattered icebergs^{4–6}. Detailed mapping of the ocean floor is the best means to resolve this issue. Although sea-floor imagery has been used to reconstruct the glacial history of the Antarctic shelf^{7–9}, little data have been collected in the Arctic Ocean because of operational constraints^{10,11}. The use of a geophysical mapping system during the submarine SCICEX expedition in 1999¹² provided the opportunity to perform such an investigation over a large portion of the Arctic Ocean. Here we analyse backscatter images and sub-bottom profiler records obtained during this expedition from depths as great as 1 kilometre. These records show multiple bedforms indicative of glacial scouring and moulding of sea floor, combined with large-scale erosion of submarine ridge crests. These distinct glaciogenic features demonstrate that immense, Antarctic-type ice shelves up to 1 kilometre thick and hundreds of kilometres long existed in the Arctic Ocean during Pleistocene glaciations.

The central Arctic Ocean contains relatively shallow areas (water depths <1,000 m; see Fig. 1) on Yermak plateau, Lomonosov ridge and Chukchi borderland—which includes Chukchi plateau, Chukchi rise and Northwind ridge. During the SCICEX-99 expedition, conducted on the nuclear-powered submarine USS *Hawkbill*, shallow sea-floor areas were targeted for mapping to detect glaciogenic bedforms. Sea-floor images (collected using a submarine-mounted 12-kHz swath bathymetry and sidescan sonar¹²) from the Chukchi borderland and the Lomonosov ridge show a variety of bedforms, including random or subparallel scours, parallel lineations, and transverse ridges. On the records from the chirp sub-bottom profiler, these bedforms are associated with planed ridge crests with rough microrelief and obvious angular unconformities cut into the stratified sediments.

Randomly oriented furrows, typically <100-m wide and up to 30-m deep, densely cover the shallowest, <400-m-deep portions of sea floor on the Chukchi borderland and adjacent continental margin (Fig. 2a). Isolated larger scours up to 700-m wide and over 10-km long occur as deep as 500 m. Even greater depths, exceeding 900 m, are attained by closely spaced, subparallel scours on the Lomonosov ridge. Sea-floor scours are known to be formed by the drift of icebergs and pack-ice ridges¹³. At present, icebergs in the Arctic Ocean have at most 50-m draughts¹⁴, whereas icebergs off Antarctica and Greenland reach depths of 500–550 m (refs 15, 16). The largest depths of gouged sea floor, extending to 850 m, have been reported from the Yermak plateau¹⁰, matching the depth of scours on the Lomonosov ridge.

Below the depth range of dense scouring, the sea floor exhibits

|| Present address: Center for Coastal Mapping, University of New Hampshire, Durham, New Hampshire 03824, USA.

coherent sets of evenly spaced, parallel, streamlined, low-relief lineations extending to 700-m depth on Chukchi borderland and to 1,000 m on Lomonosov ridge (Figs 2–4). At their shallow termination, these features are cross-cut and obscured by iceberg scours. The lineations have crest-to-crest spacing from <50 to 200 m and maximum lengths in excess of 15 km. These parallel, streamlined bedforms, or flutes, have not been observed previously in the central Arctic basin, but have been reported from surrounding glaciated shelves^{13,17–19} and around the Antarctic where they extend to 1,200-m depth^{7–9}. Flutes on Arctic shelves are relatively small, mostly under 30-m wide and 1-km long; whereas at the deep

Antarctic margin, the lineations are even larger than those discovered by SCICEX in the deep Arctic Ocean. The formation of such lineations has been explained by moulding of the soft bed by ice streams—the only known mechanisms for producing the sustained linear bedform fabric over large distances. Some fluted sites on Chukchi borderland include drumlin-like features that help to identify the direction of ice flow. Several large, almost 100-m-diameter blocks of sediment or rock with high acoustic return occur in the fluted area at the northern slope of Chukchi plateau (Fig. 2b); low-return ‘tails’ aligned with flutes indicate that these blocks were squeezed into sediment and dragged by ice.

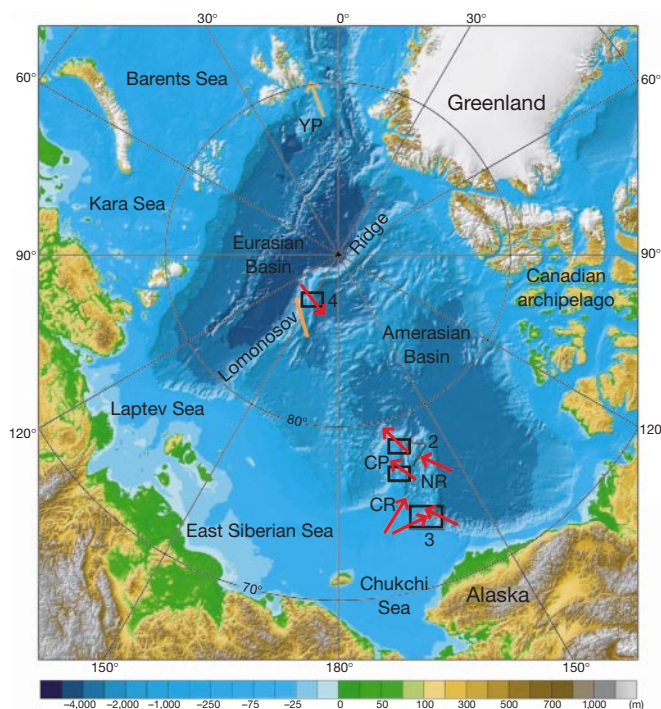


Figure 1 Map of the Arctic Ocean³⁰. Red arrows, reconstructed directions of ice-shelf flows; orange line, prevailing orientation of iceberg scours on Lomonosov ridge (this study, direction unresolved) and on Yermak plateau¹⁰. Numbered boxes show locations of

Figs 2–4. CP, Chukchi plateau; CR, Chukchi rise; NR, Northwind ridge; YP, Yermak plateau.

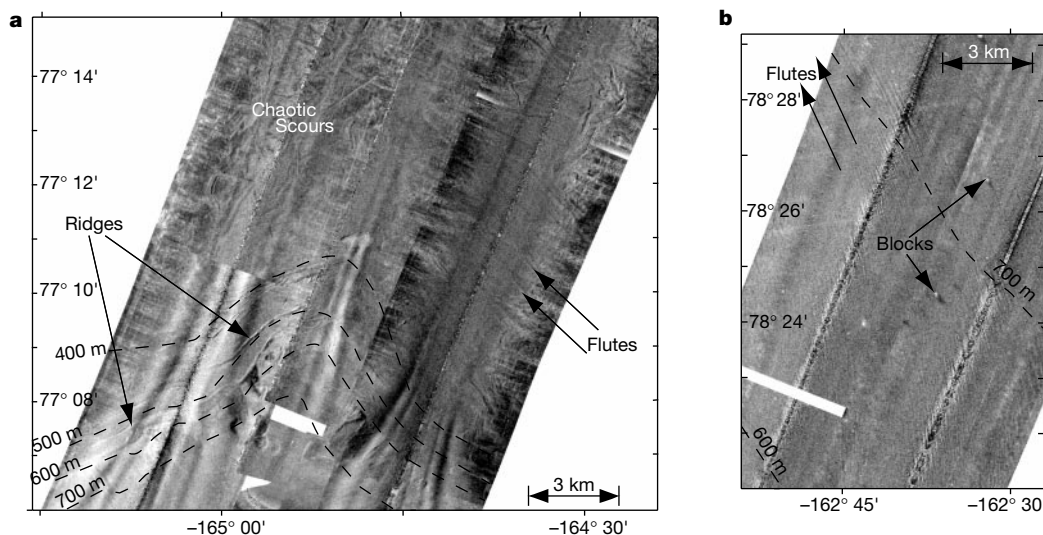


Figure 2 Swath sonar images from Chukchi plateau with overlain depth contour lines. Spatial resolution of gridded data is 16 m. **a**, Southern slope of the plateau with back-stepping ice-marginal ridges, northwest-trending flutes, and chaotic iceberg scours.

b, Northern slope of the plateau with northwest-trending flutes and scattered blocks of sediment/rock.

In several areas on Chukchi borderland, the fluted sea-floor pattern is complicated by nested sets of transverse, slightly sinuous or arcuate ridges reaching in excess of 100 m in width (Fig. 2a). These ridges parallel bathymetric contours, indicating that their formation was controlled by sea level. Such ridges are common for glaciated shelves, and have been interpreted to be formed by retreating ice-sheet margins (grounding lines)^{7,8,17–19}.

Identified bedforms from various sites can be tentatively correlated using their bathymetric position, orientation and preservation. Streamlined lineations are the best indicators of grounded ice flow^{7,8,17–19}. On Chukchi plateau, Northwind ridge, and the eastern slope of Chukchi rise, flutes consistently trend southeast–northwest (Fig. 1), with minor deviations on local topographic highs (Fig. 3). This orientation points at eastern Alaska and/or the western part of the Canadian Arctic archipelago as an ice-flow source. We suggest that the broad straits of the archipelago were the most likely outlets for major ice streams, consistent with glaciogenic features observed along the strait flanks^{20,21}. To the west, the Chukchi rise is covered by flutes that have a different orientation, indicative of ice movement from the Chukchi shelf (Figs 1, 3). These flutes do not extend to water depths >420 m, indicating thinner ice. From the overall regional pattern, we infer that a major ice shelf was propelled westwards by Alaskan/Canadian ice streams, but was deflected by the steep continental slope and/or by the Chukchi ice sheet and thus overran the Chukchi plateau in a northwest-trending direction. The general pattern of this ice movement over the Amerasian basin matches the distribution of ice-rafted debris deposited on the sea floor during glacial periods²².

The direction of ice flow across the fluted area on Lomonosov ridge is revealed in the chirp-sonar records (Figs 1, 4). Here, the crest is planed by erosion at a depth of almost 1,000 m; this planing occurs over a stretch at least 50-km wide, and a 20-m-thick lens of

acoustically transparent sediment descends from the eroded surface >100 m down the Amerasian flank of the ridge. The lens volume approximately matches the amount of sediment, with a maximum thickness of >50 m missing on top of the ridge as estimated by the projection of truncated strata. These features give a vivid picture of erosion of the crest by a single ice massif, moving north from the shelf of the Barents and Kara seas, with a debris lobe of reworked material pushed downslope on the lee side (Figs 1, 4). The dipping of the eroded surface towards the Eurasian side further shows that the ridge was a barrier to the passage of ice. The configuration of the eroded crest and the debris lobe, combined with the evenly fluted surface, indicate the sustained direction and strength of ice motion. The backstress of this motion had to be very high, at least equivalent to the ~600 kPa estimated for ice erosion on Yermak plateau³. These features suggest that the erosion on Lomonosov ridge was produced by a growing ice shelf rather than a random, abnormally large tabular iceberg, especially since the general Pleistocene iceberg drift in the Arctic Ocean was in the opposite direction²². Moreover, the fluted sea bed indicates an active grounded ice flow, not a 'bulldozing' action of icebergs. Similar large-scale erosional features occur on 500–550-m deep topographic highs on Northwind ridge, which have ~30-m thick, moraine-like, stacked lobes of acoustically transparent sediment on top. Orientation of these lobes attests to the motion of ice from the Alaskan/Canadian margin, consistent with the spatial distribution of flutes in the Chukchi region (Fig. 1).

The small ridges that are transverse to glacial lineations on Chukchi borderland probably mark the back-stepping of the ice-shelf grounding line with rising sea level during deglaciation (Fig. 2a)^{8,19}. These ridges do not occur at maximal water depths reached by flutes, which may indicate fast changes in ice-shelf profile during the initial stages of deglaciation. Iceberg scouring, which has obliterated the shallower bedforms reflecting the coherent ice

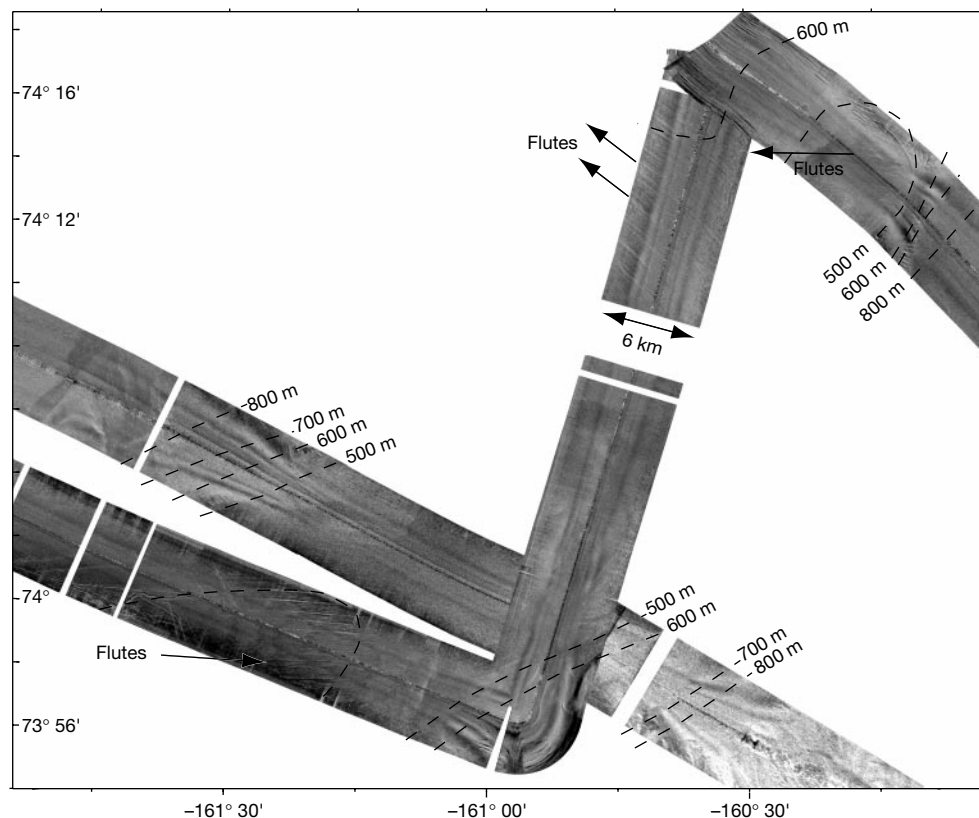


Figure 3 Swath sonar images from the eastern part of Chukchi rise with overlain depth contour lines. Spatial resolution of gridded data is 16 m. Two major sets of flutes point at the Chukchi shelf and at the Alaskan/Canadian margin as source areas. A local high

in the northeastern corner has a different orientation of flutes, presumably of younger age.

movement, represents conditions after the major ice-shelf disintegration. Subparallel scours at shallowest depths on Lomonosov ridge (750–950 m) indicate a flow of megabergs, possibly driven by a sustained current and/or jammed in an ice-bound Arctic Ocean, as suggested for similar features on Yermak plateau^{10,23}. Random scouring at yet shallower depths of <350–400 m in the Chukchi area reflects unrestricted movement of minor icebergs driven by winds and/or currents.

The stratigraphic boundary formed by glacial erosion on Lomonosov ridge (Fig. 4) has been recovered by coring²⁴. This erosion was interpreted to occur at the end of marine isotopic stage (MIS) 6, approximately 150 kyr ago. According to another, more conventional, age model for the Arctic Ocean sediments⁶, this stratigraphic level has a much older age of >600 kyr (MIS 16). The latter interpretation is consistent with an estimate of ~660 kyr ago for the timing of glacial grounding on Yermak plateau²⁵ and with the period of maximal glacial erosion of the Barents Sea between about 1,000 and 440 kyr ago²⁶.

In the absence of stratigraphic control for glaciogenic features on Chukchi borderland, we cannot conclude whether or not they were coeval with ice grounding on Lomonosov ridge. Moreover, the multitude of glaciogenic bedforms may represent a composite of several glacial events with a similar pattern of ice flow over Chukchi

borderland. Deviation of flutes on local shallows from the general orientation (Fig. 3) indicates the likelihood of multiple ice advances. However, we may be able to find direct evidence for only the thickest ice shelves; this is because thin ice would not reach the deep sea floor, and ice-flow markings in the shallowest areas would be obliterated by chaotic iceberg scouring.

An unfossiliferous, overcompacted sediment deposited before 13 kyr ago at ~400-m depth on Chukchi rise²⁷ possibly indicates ice grounding during the Last Glacial Maximum. The existence of an extensive ice shelf is corroborated by extremely low sedimentation rates (possibly a hiatus) and the absence of biogenic remains in sediments throughout the Amerasian basin during 13–20 kyr ago^{27,28}. We presume that the removal of this ice shelf facilitated large-scale surges upstream, such as a catastrophic discharge of 80,000 km³ of ice from the northwestern part of the Canadian archipelago before 10 kyr ago²¹. Understanding this and older events associated with growth and disintegration of Arctic ice shelves will help explain the behaviour of the West Antarctic Ice Sheet²⁹ and, in broader perspective, the stability of glacial climates. Ice shelves in the Arctic have large potential effects on climate by affecting the albedo and ocean–atmosphere heat exchange; this highlights a significant unresolved question—was the entire Arctic Ocean covered by a continuous ice shelf at some time during the Pleistocene? To answer this question further mapping of glaciogenic bedforms and their stratigraphic investigation are required; this could establish whether ice-shelf advances were synchronous in various parts of the Arctic basin.

Received 21 August; accepted 22 December 2000.

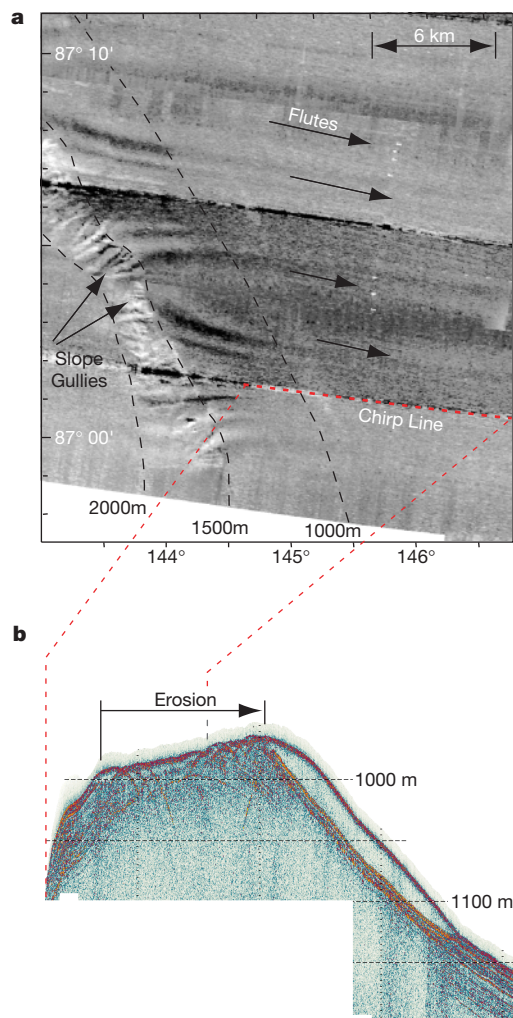


Figure 4 Swath sonar and chirp images from Lomonosov ridge. Spatial resolution of gridded sidescan data (a) is 21 m; vertical resolution of the chirp data (b) is tens of centimetres. The planed, eroded top surface of the crest is covered with ESE-trending low-relief flutes. The acoustically transparent debris lens drapes the lee (Amerasian) slope of the ridge.

- Mercer, J. H. A former ice sheet in the Arctic Ocean? *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **8**, 19–27 (1970).
- Lindstrom, D. R. & MacAyeal, D. R. Scandinavian, Siberian, and Arctic Ocean glaciation: effect of Holocene atmospheric CO₂ variations. *Science* **245**, 628–631 (1989).
- Grosswal, M. G. & Hughes, T. J. The case for an ice shelf in the Pleistocene Arctic Ocean. *Polar Geogr.* **23**, 23–54 (1999).
- Clark, D. L. Origin, nature, and world climate effect of Arctic Ocean ice cover. *Nature* **300**, 321–325 (1980).
- Phillips, R. L. & Grantz, A. Quaternary history of sea ice and paleoclimate in the Amerasia basin, Arctic Ocean, as recorded in the cyclical strata of Northwind Ridge. *Geol. Soc. Am. Bull.* **109**, 1101–1115 (1997).
- Spielhagen, R. F. et al. Arctic Ocean evidence for late Quaternary initiation of northern Eurasian ice sheets. *Geology* **25**, 783–786 (1997).
- Anderson, J. B. *Antarctic Marine Geology* (Cambridge Univ. Press, Cambridge, 1999).
- Shipp, S., Anderson, J. & Domack, E. Late Pleistocene–Holocene retreat of the West Antarctic Ice-Sheet system in the Ross Sea: Part 1—Geophysical results. *Geol. Soc. Am. Bull.* **111**, 1486–1516 (1999).
- Canals, M., Urgeles, R. & Calafat, A. M. Deep sea-floor evidence of past ice streams off the Antarctic Peninsula. *Geology* **28**, 31–34 (2000).
- Vogt, P. R., Crane, K. & Sundvor, E. Deep Pleistocene iceberg plowmarks on the Yermak Plateau: sidescan and 3.5 kHz evidence for thick calving ice fronts and a possible marine ice sheet in the Arctic Ocean. *Geology* **22**, 403–406 (1994).
- Jakobsson, M. First high-resolution chirp sonar profiles for the central Arctic Ocean reveal erosion of Lomonosov Ridge sediments. *Mar. Geol.* **158**, 111–123 (1999).
- Edwards, M. H. et al. Evidence of recent volcanic activity on the ultra-slow spreading Gakkel Ridge. *Nature* **409**, 808–812 (2001).
- Davies, T. A. et al. (eds) *Glaciated Continental Margins: an Atlas of Acoustic Images* (Chapman and Hall, London, 1997).
- Robe, R. Q. in *Dynamics of Snow and Ice Masses* (ed. Colbeck, S. C.) 211–259 (Academic, New York, 1980).
- Barnes, P. W. & Lien, R. Iceberg rework shelf sediments to 500 m off Antarctica. *Geology* **16**, 1130–1133 (1988).
- Dowdeswell, J. A., Villinger, H., Whittington, R. J. & Marienfeld, P. Iceberg scouring in Scoresby Sund and on the East Greenland continental shelf. *Mar. Geol.* **111**, 37–53 (1992).
- Josenhans, H. W. & Zevenhuizen, J. Dynamics of the Laurentide Ice Sheet in the Hudson Bay, Canada. *Mar. Geol.* **92**, 1–26 (1990).
- Solheim, A., Russwurm, L., Elverhøi, A. & Nyland-Berg, M. Glacial geomorphic features in the northern Barents Sea: direct evidence for grounded ice and implications for the pattern of deglaciation and late glacial sedimentation. *Geol. Soc. Spec. Publ.* **53**, 253–268 (1990).
- Polyak, L., Forman, S. L., Herlihy, F. A., Ivanov, G. & Krinitsky, P. Late Weichselian deglacial history of the Syvataya (Saint) Anna Trough, northern Kara Sea, Arctic Russia. *Mar. Geol.* **143**, 169–188 (1997).
- Dyke, A. S., Morris, T. F., Green, D. E. C. & England, J. Quaternary geology of Prince of Wales Island, Arctic Canada. *Geol. Surv. Can. Mem.* **433** (1992).
- Clark, C. D. & Stokes, C. R. Extent and basal characteristics of the M'Clintock Channel ice stream. *Quat. Int.* (in the press).
- Bischof, J. F. & Darby, D. A. Mid- to Late Pleistocene ice drift in the Western Arctic Ocean: evidence for a different circulation in the past. *Science* **277**, 74–78 (1997).
- Vogt, P. R., Crane, K. & Sundvor, E. Deep Pleistocene iceberg plowmarks on the Yermak Plateau: sidescan and 3.5 kHz evidence for thick calving ice fronts and a possible marine ice sheet in the Arctic Ocean (reply). *Geology* **23**, 477–478 (1995).

24. Jakobsson, M. *et al.* Pleistocene stratigraphy and paleoenvironmental variation from Lomonosov Ridge sediments, central Arctic Ocean. *Global Planet. Change* (in the press).
25. Flower, B. P. Overconsolidated section on the Yermak Plateau, Arctic Ocean: ice sheet grounding prior to ca. 660 ka? *Geology* **25**, 147–150 (1997).
26. Faleide, J. J. *et al.* Late Cenozoic evolution of the western Barents Sea–Svalbard continental margin. *Glob. Planet. Change* **12**, 53–74 (1996).
27. Darby, D. A., Bischof, J. F. & Jones, G. A. Radiocarbon chronology of depositional regimes in the western Arctic Ocean. *Deep-Sea Res. II* **44**, 1745–1757 (1997).
28. Poore, R. Z., Osterman, L., Curry, W. B. & Phillips, R. L. Late Pleistocene and Holocene meltwater events in the western Arctic Ocean. *Geology* **27**, 759–762 (1999).
29. Oppenheimer, O. Global warming and the stability of the West Antarctic Ice Sheet. *Nature* **393**, 325–332 (1998).
30. Jakobsson, M., Cherkis, N., Woodward, J., Coakley, B. J. & Macnab, R. A new grid of Arctic bathymetry: a significant resource for scientists and mapmakers. *Eos* **81**, 89, 93, 96 (2000).

Acknowledgements

We thank R. Perry (the captain), the officers and crew of the USS *Hawkbill* and the scientists and engineers who sailed during SCICEX-99; we also thank P. Johnson for assistance with figures. This work was supported by the US NSF.

Correspondence and requests for materials should be addressed to L.P.
(e-mail: polyak.1@osu.edu).

Geochemical tracing of Pacific-to-Atlantic upper-mantle flow through the Drake passage

J. A. Pearce*, P. T. Leat†, P. F. Barker† & I. L. Millar†

* Department of Earth Sciences, Cardiff University, Cardiff, CF10 3YE, UK

† British Antarctic Survey, Cambridge CB3 0ET, UK

The Earth's convecting upper mantle can be viewed as comprising three main reservoirs, beneath the Pacific, Atlantic and Indian oceans. Because of the uneven global distribution and migration of ridges and subduction zones, the surface area of the Pacific reservoir is at present contracting at about $0.6 \text{ km}^2 \text{ yr}^{-1}$, while the Atlantic and Indian reservoirs are growing at about $0.45 \text{ km}^2 \text{ yr}^{-1}$ and $0.15 \text{ km}^2 \text{ yr}^{-1}$, respectively^{1,2}. Garfunkel¹ and others have argued that there must accordingly be net mantle flow from the Pacific to the Atlantic and Indian reservoirs (in order to maintain mass balance), and Alvarez² further predicted that this flow should be restricted to the few parts of the Pacific rim (here termed 'gateways') where there are no continental roots or subduction zones that might act as barriers to shallow mantle flow. The main Pacific gateways are, according to Alvarez^{2,3}, the south-east Indian Ocean, the Caribbean Sea and the Drake passage. Here we report geochemical data which confirm that there has been some outflow of Pacific mantle into the Drake passage—but probably in response to regional tectonic constraints, rather than global mass-balance requirements. We also show that a mantle domain boundary, equivalent to the Australian–Antarctic discordance, must lie between the Drake passage and the east Scotia Sea.

Figure 1b and c shows the distribution of the $\sim 10^6 \text{ km}^2$ of new oceanic crust created between South America and Antarctica since sea-floor spreading began in the Drake passage (gateway d in Fig. 1a) about 30 Myr ago^{4,5}. Most oceanic crustal accretion took place in the Drake passage itself (about 30–10 Myr ago), the central Scotia Sea (about 20–10 Myr ago), and the east Scotia Sea (≤ 10 Myr ago). The adjacent Antarctic–Phoenix ridge was also active over much of this period (> 30 Myr ago to about 4 Myr ago). This whole 30-Myr period is thought to have been marked by westward subduction of Atlantic oceanic crust and eastward roll-back of the subducting slab.

A central question has been whether or not this roll-back is a consequence of eastward Pacific mantle flow.

Figure 1c gives the locations of samples that we have analysed. Dredges DR.62 and 63 recovered old crust from the Drake passage, and DR.7 recovered the youngest crust. DR.12 recovered old crust from the east Scotia Sea, and DR.20, 23, 107 and 118 the youngest crust. DR.3 recovered the youngest crust from the Antarctic–Phoenix ridge⁶. Our geochemical data are listed in Table 1. To assess which crust was generated from Pacific mantle, and which from Atlantic mantle, we adopt the method of isotope fingerprinting of basalts that has been used successfully to study the southeast Indian gateway (that is, the Australian–Antarctic discordance or AAD; see, for example, refs 7–9).

Figure 2a and b show two of the most effective projections for discriminating between mantle domains: plots of $^{208}\text{Pb}/^{204}\text{Pb}$ ratios versus $^{206}\text{Pb}/^{204}\text{Pb}$ ratios, and $^{143}\text{Nd}/^{144}\text{Nd}$ versus $^{206}\text{Pb}/^{204}\text{Pb}$. The Pacific domain, which extends beneath much of the Pacific Ocean, is defined here just by data from the southeast Pacific—namely, the East Pacific Rise south of the Equator^{10,11}, the Pacific–Antarctic ridge^{12–14}, and the Chile ridge^{15,16}, excluding anomalous values at the ridge–trench intersection. We have divided the South Atlantic mantle into three domains, according to geochemical coherence and proximity to hotspots. The Bouvet domain extends from Bouvet Island along the southwest Indian ridge to the west, along the South–America–Antarctic ridge to the east, and for a short distance up the Mid-Atlantic Ridge^{17–19}. The Shona and Discovery domains encompass the Mid-Atlantic Ridge²⁰ further north (Fig. 1b). It is apparent that the Pacific domain is distinct from the three Atlantic domains in its high $^{143}\text{Nd}/^{144}\text{Nd}$ and low $^{208}\text{Pb}/^{204}\text{Pb}$ for a given $^{206}\text{Pb}/^{204}\text{Pb}$, combined with its negative trend on the Nd–Pb plot and positive trend on the Pb–Pb plot. The Bouvet and Shona domains give trends similar to, but displaced from, the Pacific domain. The Discovery domain extends slightly into the Pacific domain, but forms a trend orthogonal to it.

Of our new data, the samples from the east Scotia Sea plot in the Atlantic (Bouvet) field on both diagrams (Fig. 2c, d). They have lower $^{206}\text{Pb}/^{204}\text{Pb}$ ratios than the samples from the South–America–Antarctic ridge, which reflects their greater distance from the Bouvet plume. By contrast, sample DR.7 from the Drake passage ridge plots

Table 1 Analyses from the Drake passage and surrounding region

Sample (DR.)	Location	$^{87}\text{Sr}/^{86}\text{Sr}$	$^{143}\text{Nd}/^{144}\text{Nd}$	$^{208}\text{Pb}/^{204}\text{Pb}$	$^{207}\text{Pb}/^{204}\text{Pb}$	$^{206}\text{Pb}/^{204}\text{Pb}$
62.B	DPM	0.70421	0.51286	18.992	15.621	39.028
62.153	DPM	0.70469	0.51317	18.303	15.512	37.875
63.2	DPM	0.70312	0.51294	18.850	15.584	38.465
63.20	DPM	0.70312	0.51295	18.632	15.594	38.456
3.2	APR	0.70283	0.51307	18.592	15.544	38.040
3.7	APR	0.70263	0.51310	18.371	15.526	37.888
3.4	APR	0.70350	0.51305	18.611	15.535	38.187
7.3	DPR	0.70274	0.51316	18.227	15.514	37.723
12.19	ESM	0.70309	0.51303	18.716	15.602	38.535
WX.2	ESR2	0.70289	0.51306	18.223	15.528	37.938
20.36	ESR3	0.70285	0.51308	18.115	15.493	37.733
23.1	ESR9	0.70297	0.51304	18.024	15.496	37.726
107.1	ESR1	0.70305	0.51310	17.905	15.482	37.580
118.3	ESR1	0.70325	0.51304	18.211	15.551	37.874

Isotope analyses of basic volcanic rocks from the Drake passage margin (DPM), Antarctic–Phoenix ridge (APR), Drake passage ridge (DPR), east Scotia Sea margin (ESM) and east Scotia ridge (ESR_n where *n* is the segment number²¹). Ages are 28 Myr for the DPM samples, 10 Myr for the DPR and ESM samples, and < 1 Myr for the ESR samples. Samples from the APR, DPR and DR.62.153 are MORBs, other samples from the DPM are tholeiitic (DR.63.20) and alkalic (DR.62.B and DR.63.2) ocean island basalts, samples from the ESR are MORBs with negligible to small subduction components, and the sample from the ESM is a MORB with a more significant subduction component. All analyses are new and on dredge (DR.) samples except for that from the wax-core glass (WX.20) which is published²² and included for comparison. Isotope ratios are by thermal ionization mass spectrometry²³. Over the period of this work, three analyses of NBS987 gave Sr isotope ratios of 0.710198 ± 0.000022 (2 σ) and the data were normalized to 0.710240. Five analyses of the La Jolla standard gave Nd isotope ratios of 0.511900 ± 0.000022 (2 σ) and the data were normalized to 0.511864. Pb isotope ratios were corrected for fractionation using the replicate analyses of the NBS981 standard. Sr and Nd isotope ratios have internal precisions (1 s.e.) of $\pm (3-4) \times 10^{-6}$ and $\pm (6-8) \times 10^{-6}$ respectively, and Pb isotope ratios have internal precisions of $\pm (1-2) \times 10^{-6}$ (6/4 and 7/4) and $\pm (2-4) \times 10^{-6}$ (8/4). See Table 1 in Supplementary Information for dredge depths and locations, and for the X-ray fluorescence major-element and inductively-coupled plasma mass spectrometer trace-element analyses.

within the Pacific domain, as do the samples from the Antarctic–Phoenix spreading centre. Of the four analysed samples from the Drake passage margin, sample DR.62.153, which is most like mid-ocean ridge basalt (MORB), is almost identical to the sample from the Drake passage ridge. The three other samples, which are more like ocean island basalt (OIB), plot within the Bouvet domain near Bouvet Island itself. The OIB samples are unlikely to have had asthenospheric sources as they erupted more than 1,000 km from any plume. The probable explanation is that the rifting stage of the opening of the Drake passage tapped lithosphere that had been modified by the Bouvet plume before Gondwana breakup. This was followed by a drifting stage, which tapped asthenosphere of Pacific provenance. The basalt from DR.12 also has a much higher $^{206}\text{Pb}/^{204}\text{Pb}$ ratio than is likely for an asthenospheric source, and may also owe its composition to enriched lithospheric mantle tapped during the initial stages of the opening of the east Scotia Sea. We note that DR.12 also carries a subduction component, as indicated by its elevated Th/Ta ratio relative to MORB (3.3 compared to < 2 for the other Scotia Sea samples), but is plotted because mixing lines to subducted crust and sediment do not cross the Pacific–Atlantic discriminant boundary. We also note that the Drake passage MORB samples in fact plot in the overlap field of the Pacific and Discovery plume domains (Fig. 2a, b). We interpret them as Pacific, first on statistical grounds (probability density functions show the Discovery interpretation to have $< 10\%$ probability), and second because the data from the east Scotia Sea indicate that the ambient Atlantic mantle belongs to the Bouvet,

rather than Discovery, domain.

Barker's⁵ reconstruction of the Scotia Sea region may provide clues to how these mantle domains interacted. We assume that before the opening of the Drake passage (Fig. 3a), the Andean margin separated mantle of Pacific and Atlantic provenance. During the first stage of the opening (Fig. 3a, b), the South Atlantic mantle would have been restricted from entering the Drake passage—according to the Alvarez² hypothesis—because of the bordering continental fragments and subducting slabs. The composition of sample DR.62.153 confirms that the first mantle asthenosphere to melt was of Pacific provenance. In the second stage of opening (Fig. 3b, c), the barrier between the Pacific and Atlantic upper-mantle reservoirs may have first been breached during spreading in the central Scotia Sea. This would have allowed Pacific and Atlantic mantle to mix at shallow depths, although we have at present no data to constrain this. We do know, however, that the Drake passage continued to be fed by mantle of Pacific provenance 10 Myr ago (sample D.7.3). During the third stage (Fig. 3c, d), the samples from the east Scotia Sea show that rapid retreat of the south Sandwich trench was fed not by eastward flow of Pacific mantle but by flow of South Atlantic mantle around the retreating trench. In consequence, the present domain boundary probably lies in or adjacent to the central Scotia Sea.

Thus the data suggest that the Pacific mantle entered the Drake passage about 30 Myr ago, and continued to enter as the passage opened. Once spreading stopped, about 10 Myr ago, there is evidence not for continued eastward Pacific mantle flow, but instead

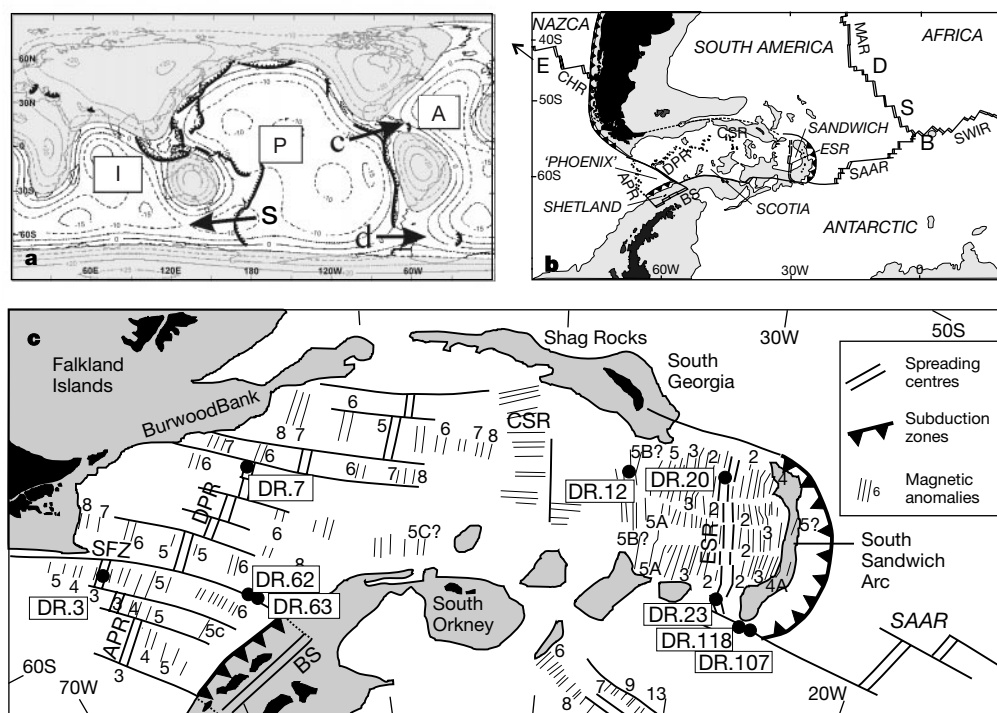


Figure 1 Location and tectonic setting of the Drake passage and of the samples that we analysed. **a**, Modelled perturbation in crustal thickness (model M84C of ref. 30 contoured at 5-km intervals and shaded at > 5 km), highlighting the three principal upper-mantle reservoirs (A, Atlantic; I, Indian; and P, Pacific) and the main links between them (s, southeast Indian Ocean; c, Caribbean; and d, Drake passage). We term these links 'gateways' as they are mantle analogues to the 'gateways' that permit seawater circulation through continental masses. **b**, Regional tectonic setting of the Drake passage/Scotia Sea 'gateway' (d in **a**). Spreading centres are depicted by double lines (continuous for active, dashed for extinct), trenches by toothed lines and transform faults by single lines. Plates are named. Subaerial areas of continents are dark grey, submerged areas at < 3 -km depth are pale grey. Key to spreading centres: MAR, Mid-Atlantic Ridge; SWIR,

southwest Indian ridge; SAAR, South-America–Antarctic ridge; CHR, Chile ridge; APR, Phoenix–Antarctic ridge (extinct); DPR, Drake passage ridge (extinct); CSR, central Scotia ridge (extinct); and ESR, east Scotia Sea ridge. BS, Bransfield strait. Plume locations are: B, Bouvet; S, Shona; D, Discovery; and E, Easter. **c**, Detailed map (Lambert conical orthomorphic projection) of the Drake passage/Scotia Sea 'gateway'⁴ showing the areas of new oceanic crust (at minimum, those with magnetic anomalies) created since the split of South America and Antarctica, together with the locations of the samples (DR.x) analysed here. Numbers refer to identified magnetic anomalies. Known continental and arc crust is shown as black above sea level and grey below sea level. SFZ, Shackleton fracture zone.

for flow of Atlantic mantle into the east Scotia Sea. This is consistent with shear-wave splitting experiments, which provide positive evidence for present-day eastward flow through the Caribbean gateway^{21,22} but not for eastward flow through the Drake passage²³. It is also consistent with geoid anomaly maps which show a high over the South Atlantic hotspots but a low over the central, western Atlantic²⁴, potentially providing a positive pressure gradient for eastward flow through the Caribbean—but not for flow through the Drake passage. The narrowness of the subducting slab may have also been a factor^{25,26}, enabling the negative pressure (suction) at the two slab edges to induce northward and southward flow of Atlantic mantle into the east Scotia Sea (Fig. 3d). A wider slab would have been more likely to induce eastward corner flow, which would have favoured the influx of Pacific mantle. Our inference is therefore that the opening of the Drake passage has allowed a limited volume of mantle to flow out of the Pacific reservoir, but that it has not provided the main gateway feeding the Mid-Atlantic Ridge that Alvarez² proposes. Indeed, it is more likely that the regional extensional tectonics which preceded the opening of the gateway led to the outflow of Pacific mantle and that, once the gateway opened, global pressure gradients acted to oppose that flow.

The volume of mantle leaving the Pacific reservoir through the Drake passage must be sufficient to account for an increase in Atlantic surface area of $0.02 \pm 0.005 \text{ km}^2 \text{ yr}^{-1}$ (the average area of new ocean floor created per year in the Drake passage over the past 30 Myr). This compares with about $0.05 \text{ km}^2 \text{ yr}^{-1}$ for the southeast Indian gateway over the same period⁸. These two gateways can

therefore account for little more than 10% of the $0.6 \text{ km}^2 \text{ yr}^{-1}$ shrinkage of the Pacific reservoir over this period. However, Alvarez does not consider the extent to which plume-related mantle upflow might counter the subduction-related oceanic lithospheric downflow that leads to the shrinkage of the reservoir (see, for example, ref. 24). Sleep's estimate²⁷, based on swell uplift, gives a combined Indian and Atlantic plume upflow of about $24 \text{ km}^3 \text{ yr}^{-1}$ for a plume excess temperature of 300°C . In this case, the plume flux can account for only about $0.1 \text{ km}^2 \text{ yr}^{-1}$ of the $0.6 \text{ km}^2 \text{ yr}^{-1}$ shrinkage of the Pacific reservoir, assuming that the plume feeds a 250-km-thick asthenosphere layer. In contrast, Phipps Morgan *et al.*²⁸ propose that the global plume upflow balances the global lithospheric downflow of about $240 \text{ km}^3 \text{ yr}^{-1}$. On this basis, and using Sleep's estimate that about 40% of the global plume flux is non-Pacific, we obtain a higher Indian and Atlantic plume upflow of about $100 \text{ km}^3 \text{ yr}^{-1}$, which would account for much ($\sim 0.4 \text{ km}^2 \text{ yr}^{-1}$) of the shrinkage. Thus, if the estimate by Sleep²⁷ is more realistic, then neither plumes, nor mantle outflow through the Drake passage and southeast Indian gateways, are sufficient to explain the shrinkage of the Pacific reservoir, and mass balance must be achieved through other gateways such as the Caribbean, or by other processes such as flow of sub-continental mantle asthenosphere. However, if the estimate by Phipps Morgan *et al.*²⁸ is more realistic, then other mantle gateways need make only relatively minor contributions to upper-mantle mass balance.

Our data achieve the following results. They extend the area of mantle of known Pacific provenance far to the southeast of its

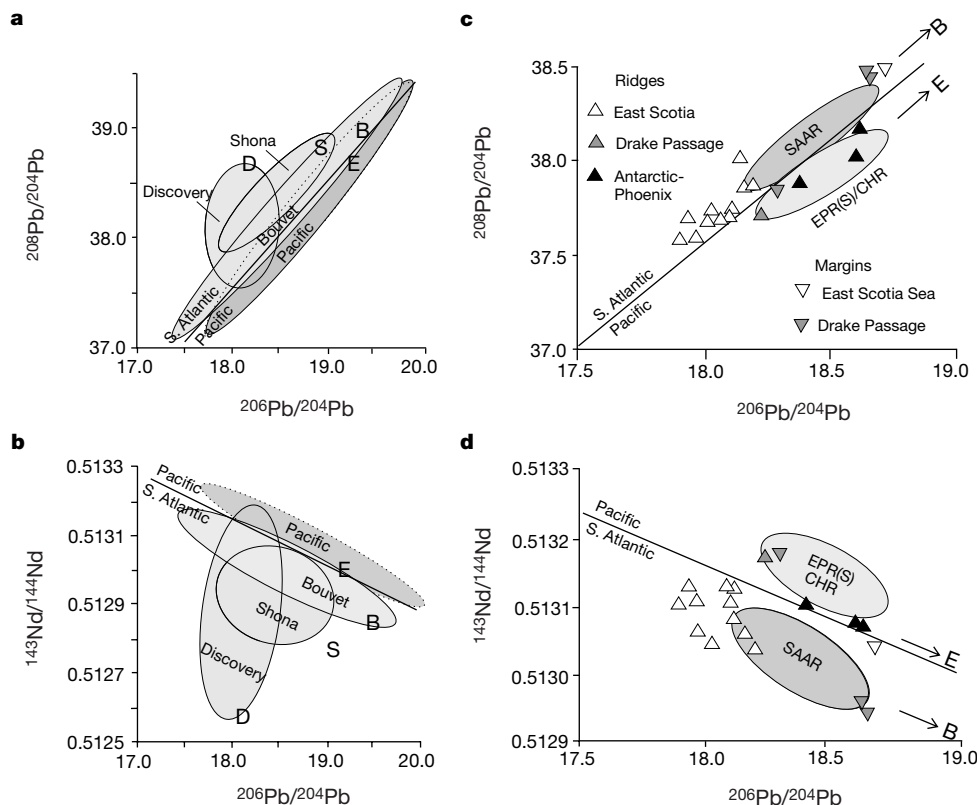


Figure 2 Plots of $^{208}\text{Pb}/^{204}\text{Pb}$ versus $^{206}\text{Pb}/^{204}\text{Pb}$ and $^{143}\text{Nd}/^{144}\text{Nd}$ versus $^{206}\text{Pb}/^{204}\text{Pb}$, for discriminating between Pacific and South Atlantic MORB and fingerprinting Drake passage samples. **a, b**, Dispersions of the mantle domains, represented as 95% probability ellipses, with the discriminant boundary between Pacific and South Atlantic mantle drawn by eye. The latter is subdivided into domains according to proximity to the Bouvet, Shona and Discovery plumes. The nearest plume in the southeast Pacific is Easter Island. Average end-members for these plumes are shown by initial letters; that is, B, S, D and E respectively. **c, d**, Expanded-scale diagrams for the samples from the Drake

passage and the east Scotia Sea and the nearest Pacific and Atlantic ridges. The east Scotia Sea data set is augmented by published analyses²⁹ of the most MORB-like glasses from segment 2 (north of DR.20) and our additional, unpublished data from DR.107 and DR.118. Acronyms refer to the various ridges in Fig. 1. The MORB sample from the Drake passage margin (DR.62.153) plots close to, and is grouped with, the Drake passage ridge sample. The diagrams show a clear distinction between the Drake passage and the east Scotia Sea, and suggest the former to be of Pacific provenance and the latter of Atlantic provenance.

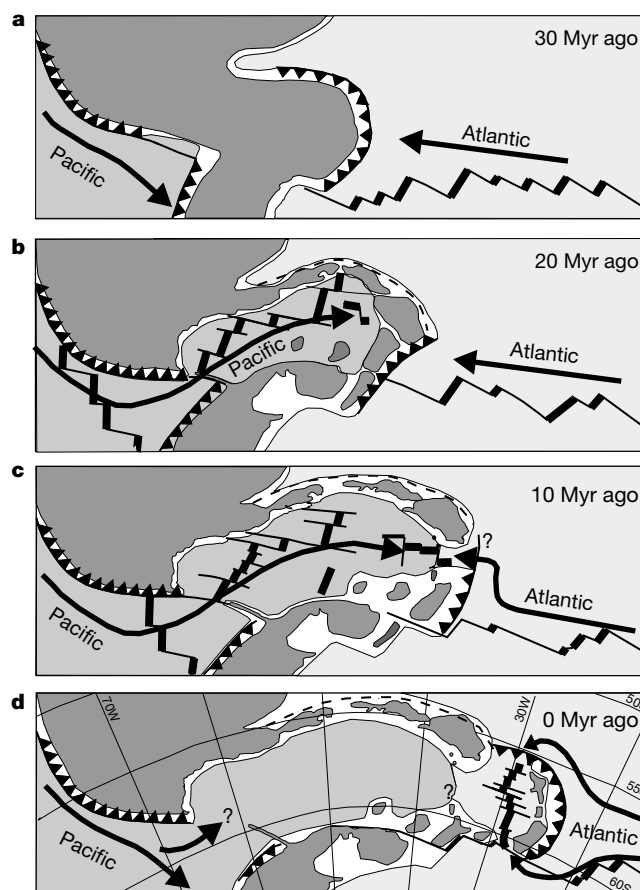


Figure 3 Model for the mantle dynamics of the Drake passage/Scotia Sea 'gateway' and surrounding region. The shaded areas depict sub-continental mantle (deep grey), sub-oceanic mantle of Pacific provenance (mid-grey), sub-oceanic mantle of Atlantic provenance (pale grey), and undetermined (white). Arrows depict inferred directions of mantle flow. Tectonic reconstructions are those of Barker⁵. **a**, At 30 Myr ago, western South America is part of a continuous Andean plate margin and westward subduction has just initiated in the Scotia region. At this point, continental lithosphere and subduction zones separate shallow, convecting mantle of Pacific and Atlantic provenance. **b**, By

20 Myr ago, the Drake passage is opening, allowing Pacific mantle to enter from the west. Influx of Atlantic mantle from the east is prevented by the surrounding continental fragments and subduction zones. **c**, By 10 Myr ago, the Drake passage has continued to open and the central Scotia Sea has opened. This provides the opportunity for the two mantle domains to interact, but as yet there are no data to constrain this. **d**, By 0 Myr ago, the east Scotia Sea has opened, accompanied by influx of Atlantic mantle from the east, driven by Bouvet plume and subduction-related pressure gradients. (Lambert conical orthomorphic projection; South America fixed for the reconstruction).

previously known geographic limits^{12–14}. They demonstrate that isotope ratios are, as in the case of the AAD, viable tracers for studying Pacific mantle flow into the southernmost Atlantic. They indicate that eastward Pacific mantle flow did take place while the Drake passage was actively extending. However, they do not support the Alvarez hypothesis that the Drake passage is at present a gateway for mantle flow from the Pacific to the Atlantic, instead favouring westward, plume- and subduction-driven mantle flow as the stronger driving force. Finally, they suggest that a mantle domain boundary comparable to the AAD must lie somewhere in the region of the central Scotia Sea. □

Received 9 May 2000; accepted 15 January 2001.

- Garfunkel, Z. Growth, shrinking, and long-term evolution of plates and their implications for the flow pattern in the mantle. *J. Geophys. Res.* **80**, 4425–4432 (1975).
- Alvarez, W. Geological evidence for the geographical pattern of mantle return flow and the driving mechanism of plate tectonics. *J. Geophys. Res.* **87**, 6697–6710 (1982).
- Alvarez, W. Geologic evidence for the plate driving mechanism: the continental undertow hypothesis and the Australian–Antarctic Discordance. *Tectonics* **5**, 1213–1229 (1990).
- Barker, P. F. & Burrell, J. The opening of Drake Passage. *Mar. Geol.* **25**, 15–34 (1977).
- Barker, P. F. in *Backarc Basins: Tectonics and Magmatism* (ed. Taylor, B.) 281–314 (Plenum, New York, 1995).
- Saunders, A. D., Tarney, J., Weaver, S. D. & Barker, P. F. in *Antarctic Geoscience* (ed. Craddock, C.) 213–222 (Univ. Wisconsin Press, Madison, 1982).
- Klein, E. M., Langmuir, C. H., Zindler, A., Staudigel, H. & Hamelin, R. Isotope evidence of a mantle convection boundary at the Australian–Antarctic Discordance. *Nature* **333**, 623–629 (1988).

- Pyle, D. G., Christie, D. M., Mahoney, J. J. & Duncan, R. A. Geochemistry and geochronology of ancient southeast Indian and southwest Pacific seafloor. *J. Geophys. Res.* **100**, 22261–22282 (1995).
- Christie, D. M., West, B. P., Pyle, D. G. & Hanan, B. B. Chaotic topography, mantle flow and mantle migration in the Australian–Antarctic discordance. *Nature* **394**, 637–644 (1998).
- Mahoney, J. J. *et al.* Isotope and trace-element characteristics of a super-fast spreading ridge – East Pacific Rise, 13–23°S. *Earth Planet. Sci. Lett.* **121**, 173–193 (1994).
- Bach, W., Hegner, E., Erzinger, J. & Satir, M. Chemical and isotopic variations along the superfast spreading East Pacific Rise from 6 to 30°S. *Contrib. Mineral. Petrol.* **116**, 365–380 (1994).
- Castillo, P. R., Natland, J. H., Niu, Y. & Lonsdale, P. F. Sr, Nd and Pb isotopic variation along the Pacific–Antarctic rise crest, 53–57°S: implications for the composition and dynamics of the South Pacific upper mantle. *Earth Planet. Sci. Lett.* **154**, 109–125 (1998).
- Vlastélic, I. *et al.* Large-scale chemical and thermal division of the Pacific mantle. *Nature* **399**, 345–350 (1999).
- Ferguson, E. M. & Klein, E. M. Fresh basalts from the Pacific–Antarctic Ridge extend the Pacific geochemical province. *Nature* **366**, 330–333 (1993).
- Klein, E. M. & Karsten, J. L. Ocean-ridge basalts with convergent-margin geochemical affinities from the Chile Ridge. *Nature* **374**, 52–57 (1995).
- Bach, W. *et al.* Unusually large Nb–Ta depletions in North Chile ridge basalts at 36°50′ to 38°56′: major element, trace element and isotopic data. *Earth Planet. Sci. Lett.* **142**, 223–240 (1996).
- le Roex, A. P. *et al.* Petrology and geochemistry of basalts from the American–Antarctic Ridge, Southern Ocean: implications for the westward influence of the Bouvet mantle plume. *Contrib. Mineral. Petrol.* **90**, 367–380 (1985).
- Mahoney, J. J., le Roex, A. P., Peng, Z., Fisher, R. L. & Natland, J. H. Southwestern limits of Indian Ocean Ridge mantle and the origin of low ²⁰⁶Pb/²⁰⁸Pb mid-ocean ridge basalt: isotope systematics of the Central Southwest Indian Ridge (17°–50°E). *J. Geophys. Res.* **97**, 19771–19790 (1992).
- Kurz, M. D., le Roex, A. & Dick, H. J. B. Isotope geochemistry of the oceanic mantle near the Bouvet triple junction. *Geochim. Cosmochim. Acta* **62**, 841–852 (1998).
- Douglass, J., Schilling, J.-G. & Fontignie, D. Plume-ridge interactions of the Discovery and Shona mantle plumes with the southern mid-Atlantic Ridge (40–55°S). *J. Geophys. Res.* **104**, 2941–2962 (1999).

21. Russo, R. M. & Silver, P. G. Trench-parallel flow beneath the Nazca plate from seismic anisotropy. *Science* **263**, 1105–1111 (1994).
22. Russo, R. M., Silver, P. G., Franke, M., Ambeh, W. B. & James, D. E. Shear-wave splitting in northeast Venezuela, Trinidad and the eastern Caribbean. *Phys. Earth Planet. Inter.* **95**, 251–275 (1996).
23. Helffrich, G., Wiens, D., Barrientos, S. & Vera, E. Mantle flow around South America through the Drake Passage: teleseismic shear wave splitting results from the SEPA. *J. Conf. Abs.* **4**, 844 (1999).
24. Phipps Morgan, J., Morgan, W. J., Zhang, Y.-S. & Smith, W. H. F. Observational hints for a plume-fed sub-oceanic asthenosphere and its role in mantle convection. *J. Geophys. Res.* **100**, 12753–12767 (1995).
25. Dvorkin, J., Nur, A., Mavko, G. & Ben-Avraham, Z. Narrow subducting slabs and the origin of backarc basalts. *Tectonophysics* **227**, 63–79 (1993).
26. Livermore, R., Cunningham, A., Vanneste, L. & Larter, R. Subduction influence on magma supply at the East Scotia Ridge. *Earth Planet. Sci. Lett.* **150**, 261–275 (1997).
27. Sleep, N. H. Hotspots and mantle plumes: some phenomenology. *J. Geophys. Res.* **95**, 6715–6736 (1990).
28. Phipps Morgan, J. & Morgan, W. J. Two-stage melting and the geochemical evolution of the mantle: a recipe for mantle plum-pudding. *Earth Planet. Sci. Lett.* **170**, 215–239 (1999).
29. Leat, P. T., Livermore, R. A., Millar, I. L. & Pearce, J. A. Magma supply in back-arc spreading centre segment E2, East Scotia Ridge. *J. Petrol.* **41**, 845–866 (2000).
30. Woodhouse, J. H. & Dziewonski, A. M. Mapping the upper mantle: three-dimensional modeling of Earth structure by inversion of seismic waveforms. *J. Geophys. Res.* **89**, 5953–5986 (1984).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank C. Ottley for analytical assistance; A. Saunders for providing some of the Scotia Sea samples; J. Mahoney, J. Phipps Morgan, G. Helffrich, G. Milne and M. Bott for discussions; and D. Christie for comments on the manuscript. This work was supported by the NERC.

Correspondence and requests for materials should be addressed to J.A.P. (e-mail: PearceJA@cardiff.ac.uk).

An exceptionally preserved vermiform mollusc from the Silurian of England

Mark D. Sutton*, Derek E. G. Briggs†, David J. Siveter‡ & Derek J. Siveter*§

* Earth Sciences Department, University of Oxford, South Parks Road, Oxford OX1 3PS, UK

† Department of Earth Sciences, Wills Memorial Building, Queen's Road, Bristol BS8 1RJ, UK

‡ Department of Geology, Bennett Building, University of Leicester, University Road, Leicester LE1 7RH, UK

§ Geological Collections, University Museum of Natural History, Oxford OX1 3PW, UK

Studies of the origin and radiation of the molluscs have yet to resolve many issues regarding their nearest relatives, phylogeny and ancestral characters^{1–11}. The Polyplacophora (chitons) and the Aplacophora are widely interpreted as the most primitive extant molluscs^{2,3,9,10}, but Lower Palaeozoic fossils of the former lack soft parts, and the latter were hitherto unrecognized as fossils. The Herefordshire Lagerstätte¹² is a Silurian (about 425 Myr BP) deposit that preserves a marine biota in remarkable three-dimensional detail. The external surface of even non-biomineralized cuticle was preserved by entombment in volcanic ash, subsequent incorporation into concretions, and infilling of the fossils with sparry calcite¹³. Here we describe, from this deposit, a complete vermiform mollusc, which we interpret as a plated aplacophoran. Serial grinding at intervals of tens of micrometres, combined with computer-based reconstruction methods, renders the fossils in the round¹⁴.

Phylum Mollusca

Acaenoplax Sutton, Briggs, Siveter and Siveter gen. nov.

Acaenoplax hayae Sutton, Briggs, Siveter and Siveter sp. nov.

1996 Type 1 polychaete worm; Briggs *et al.*¹², p. 249, fig. 1f.

2000 Polychaete worm; Orr *et al.*¹³, fig. 1c, g.

Etymology. Generic name from *akaina* (Greek, meaning spine or thorn) and *plax* (Greek, meaning plate), gender feminine; specific name after Juliet Hay.

Holotype. Oxford University Museum of Natural History; OUM 29528 (Fig. 2a, b).

Other material. Twenty-two specimens (OUM), five partially reconstructed in three dimensions.

Stratigraphy and locality. Wenlock Series, Silurian; Herefordshire, England.

Diagnosis. The body is vermiform, bearing seven isolated dorsal valves. The anterior-most valve is cap-like, the remainder are saddle-like with submedial apices ranging in outline from transversely oval to subcircular, except the posterior-most, which is subrectangular. The valves are smooth except for a band of short spines fringing the posterior half of valves 3–6; the posterior margins are non-crenulate. The mouth is terminal, surrounded by a disc-like structure. Regularly spaced short, fleshy ridges, 18 in number, bearing arrays of long spines traverse the dorsal half of the body posterior of valve 2. Correspondingly spaced, paired oblique rows of fleshy 'lobes' converge posteriorly on a median ventral groove. A posterior cavity is encased by the seventh dorsal valve and a shorter corresponding ventral valve, with posteriorly emerging stubby protuberances.

Description. Specimens are 30–40 mm long and up to 4–5 mm wide, and are normally slightly flattened dorso-ventrally and taper towards the anterior and posterior. The body was evidently flexible, particularly dorso-ventrally (for example see Fig. 1b, c, h).

The head terminates in a disc, which apparently surrounds an oral opening (Fig. 1f). The body bears seven valves, which are now calcite but assumed to have been originally aragonite. Valve 1 (Figs 1g, 2a and 3, left) is smooth and suboval in outline, and overlies the oral disc. Valves 2–6 (Fig. 3, left) are saddle-like, with a characteristic 'pinched' submedial raised region (Fig. 1c, h, k). They are suboval in outline with straighter anterior margins, the size increasing from valves 2–4 and diminishing to valve 6 (Fig. 3, left). The dorsal surface of the valves is smooth, except for a band of short spines that fringe the posterior half of valves 3–6 (Figs 1k and 3, left). The posterior extremity of the trunk is covered by the seventh dorsal and a corresponding ventral valve (Fig. 1d, e, i). The former has the same saddle-like form as valves 2–6, but is longer and more rectangular in outline. The ventral valve is shorter and convex ventrally with concave lateral facets. Growth lines in the dorsal valve (Fig. 1d) show that it grew by marginal accretion, and confirm that the dorsal and ventral valves are discrete structures.

Regularly spaced short, fleshy ridges ($n = 18$) (Fig. 3, left) traverse the dorsal half of the body behind valve 2 (Figs 1c, i, k and 3, left). The posterior surfaces of these ridges bear large spines (Fig. 1i), which are directed backwards at a high angle to the trunk. Where a valve overlies a ridge (Fig. 3, left) the spines are not developed.

The cuticle is wrinkled transversely, a feature that is most evident between valves 1 and 2 (Figs 1f–h and 2b). In places the wrinkles develop low subconical swellings or projections (for example see Fig. 1b, dorsal surface) and, except on the ventral surfaces, those posterior to valve 2 bear fine spines (Fig. 2d).

The ventral half of the trunk bears paired subconical to paddle-shaped fleshy lobes that are arranged in oblique rows converging posteriorly (Figs 1a and 3, right). Their spacing is co-incident with the dorsal ridges (except where the latter are absent in the anterior region). The largest lobe in each row is at the anterior, lying immediately in front of a ridge, and they diminish in size towards the posterior. The number of lobes in a row varies from eight in the anterior part of the trunk, and reduces to four posterior of ridge 7. The lobes bear fine spines (Fig. 2d). A shallow ventral groove (Figs 1e, j and 3, right) originates in a position between valves 1 and 2, and becomes more pronounced towards the posterior. The

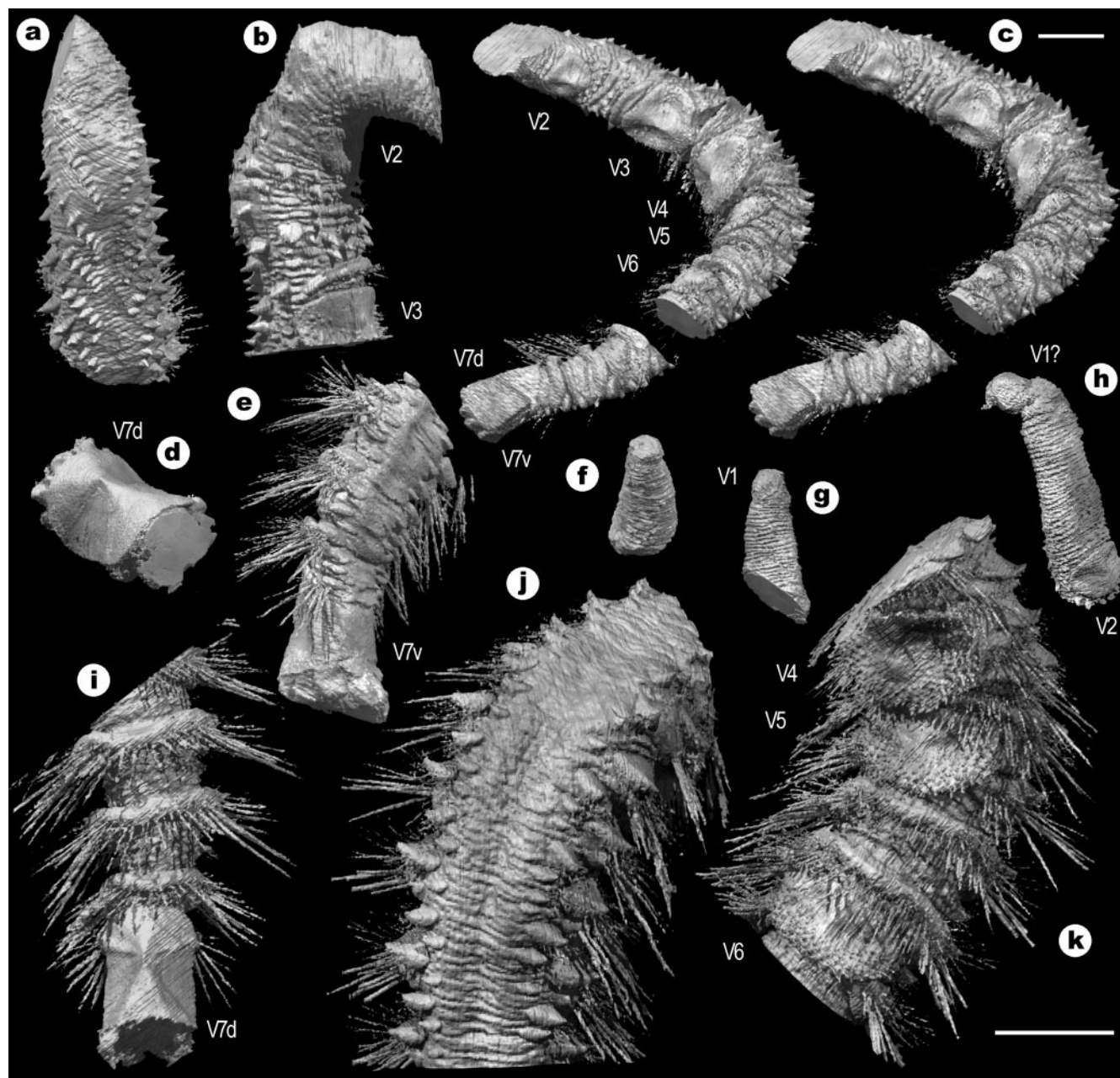


Figure 1 Reconstructed specimens of *Acaenoplax hayae*. See Fig. 3 for valve labels (V1–V7). **a**, Antero-medial fragment of OUM 29529 (ventral view) **b**, Antero-medial fragment of OUM 29530 (lateral view). **c**, Stereo pair of assembled fragments of OUM 29529 (dorsal and lateral view). **d**, Dorsal posterior valve of OUM 29531 (dorsal and lateral view, looking towards the posterior). **e**, Posterior fragment of OUM 29529 (ventral and lateral view). **f**, Anterior fragment of OUM 29530 (ventral and anterior view) **g**, As in **f**, but with dorsal

and lateral view. **h**, Anterior fragment of OUM 29532 (Dorsal and lateral view). **i**, Posterior fragment of OUM 29532 (dorsal view). **j**, Postero-medial fragment of OUM 29529 (ventral view). **k**, Same fragment as in **j**, except with a dorsal view. Scale bar: **a**, **b**, **d–k**, 2 mm (bottom right); **c**, 2 mm (top right). Anterior is to the top of the page, except in **d**. See Supplementary Information for rotating animations of these specimens.

posterior valves surround a posteriorly facing cavity (Fig. 2c), which accommodates about 12 short projections. Although the details are poorly resolved, it is probable that these represent gills. The only internal structure that is occasionally preserved is tube-like, and extends along the length of the body (Fig. 2b); it presumably represents the gut.

Affinities and evolutionary significance. *Acaenoplax* is closely related to three penecontemporaneous ‘polyplacophoran’ genera from Gotland, *Heloplax*, *Enetoplax* and *Arctoplax*, which are known from isolated valves¹⁵. These differ only in detail from *Acaenoplax* valves, and were originally aragonitic in composition¹⁶. Similarities with the polychaete annelids are convergent and a number of

characters exclude *Acaenoplax* from this group (such as marginally accreting aragonitic valves; posterior cavity with protruberances; lobes and spines that cannot be homologized with parapodia and chaetae). A molluscan affinity is strongly implied by the polyplacophoran-like dorsal scleritome, flexible body, and spines that can be homologized with the cuticular spicules of polyplacophorans and aplacophorans. Polyplacophorans have eight not seven dorsal valves, but the gap between valves 6 and 7 in *Acaenoplax* suggests that the equivalent to the 7th valve of modern chitons is not expressed. Aplacophoran-like characters include the posterior cavity with protuberances (identified as a mantle cavity with gills), the vermiform body, and the apparent absence of a foot (a

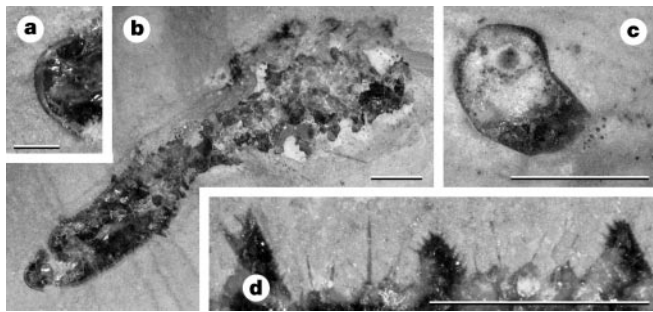


Figure 2 *Acaenoplax hayae*. **a**, Holotype OUM 29528a, detail of anterior showing V1. **b**, as in **a**, but showing the entire specimen. **c**, OUM 29532, oblique section through posterior valves and cavity. **d**, OUM 29533a, lateral cuticle from mid-trunk showing lobes, wrinkles and fine spines. Scale bar: **a**, 0.5 mm; **b–d**, 2 mm.

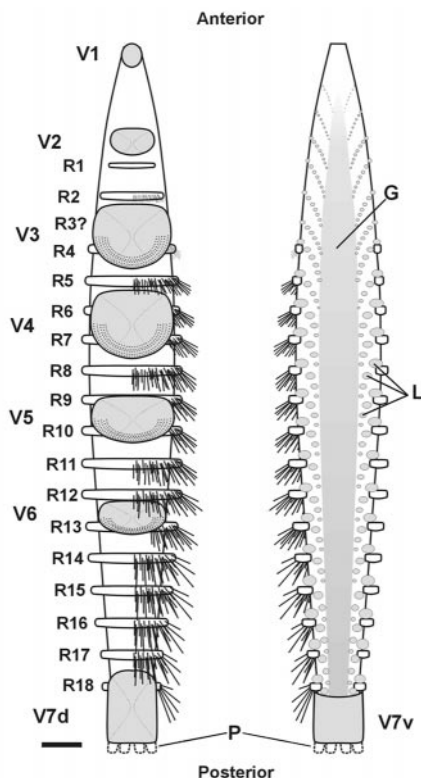


Figure 3 Idealized reconstruction of *A. hayae*. Left, dorsal view, with ridges (R1–R18) and dorsal valves (V1–V6, V7d), some bearing spines (concentric rows of dots). Right, ventral view, with ventral groove (G), lobes (L) and posterior ventral valve (V7v). Dashed outlines indicate approximation to the posterior protruberances (P). Spines arising from ridges are shown on one side only for clarity. Scale bar, ~2 mm.

character shared with the caudofoveates, or Chaetodermomorpha. This combination of molluscan characters is not found in extant taxa, but the most parsimonious placement of the new genus, in both of the two competing models of molluscan phylogeny^{3,9,10}, is as an aplacophoran sister group to the caudofoveates. Molluscs are widely believed to be primitively pseudo-metameric^{3,9,17}, but the well-developed serial repetition of *Acaenoplax* is at a higher frequency than that of other known taxa, and may provide insight on the primitive molluscan state. □

Methods

We produced three-dimensional reconstructions by serial grinding, followed by assembly of images into volume models and the calculation of isosurfaces¹⁴. Isosurfaces are rendered by ray-tracing with a primary light source above and left of the camera. Models in Fig. 1

were 'virtually prepared' by removal of objects not continuous with the fossil and of structures evident through the sparry calcite at levels below that of the exposed surface.

Received 21 July; accepted 19 December 2000.

1. Bengtson, S. The cap-shaped Cambrian fossil *Maikhanella* and the relationship between coeloclerithophorans and molluscs. *Lethaia* **25**, 401–420 (1992).
2. Conway Morris, S. & Peel, J. S. Articulated halkieriids from the Lower Cambrian of North Greenland and their role in early protostome evolution. *Phil. Trans R. Soc. Lond. B* **347**, 305–358 (1995).
3. Ivanov, D. L. in *Origin and Evolutionary Radiation of the Mollusca* (ed. Taylor, J. D.) 59–65 (Oxford Univ. Press, 1996).
4. Lindberg, D. R. in *Origin and Evolutionary Radiation of the Mollusca* (ed. Taylor, J. D.) 67–75 (Oxford Univ. Press, 1996).
5. Peel, J. S. Functional morphology, evolution and systematics of early Palaeozoic univalved molluscs. *Bull. Grönlands Geol. Unders.* **161**, 1–116 (1991).
6. Rosenberg, G. et al. Ribosomal RNA phylogeny of selected major clades in the Mollusca. *J. Mollus. Stud.* **63**, 301–309 (1997).
7. Runnegar, B. in *Origin and Evolutionary Radiation of the Mollusca* (ed. Taylor, J. D.) 77–87 (Oxford Univ. Press, 1996).
8. Runnegar, B. N. & Pojeta, J. in *The Mollusca* Vol. 10 (eds Trueman, E. R. & Clarke, M. R.) 1–57 (Academic, Orlando, 1985).
9. Salvini-Plawen, L. v. & Steiner, G. in *Origin and Evolutionary Radiation of the Mollusca* (ed. Taylor, J. D.) 29–51 (Oxford Univ. Press, 1996).
10. Scheltema, A. H. in *Origin and Evolutionary Radiation of the Mollusca* (ed. Taylor, J. D.) 53–58 (Oxford Univ. Press, 1996).
11. Winnepeinckx, B., Backeljau, T. & De-Wachter, R. Investigation of molluscan phylogeny on the basis of 18S rRNA sequences. *Mol. Biol. Evol.* **13**, 1306–1317 (1996).
12. Briggs, D. E. G., Siveter, David J. & Siveter, Derek J. Soft-bodied fossils from a Silurian volcanoclastic deposit. *Nature* **382**, 248–250 (1996).
13. Orr, P. J., Briggs, D. E. G., Siveter, David J. & Siveter, Derek J. Three-dimensional preservation of a non-biom mineralized arthropod in concretions in Silurian volcanoclastic rocks from Herefordshire, England. *J. Geol. Soc. Lond.* **157**, 173–186 (2000).
14. Sutton, M. D., Briggs, D. E. G., Siveter, David J. & Siveter, Derek J. Methodologies for the visualization and reconstruction of three-dimensional fossils from the Silurian Herefordshire lagerstätte. *Paleontol. Electronica* (in the press).
15. Cherns, L. Silurian polyplacophoran molluscs from Gotland, Sweden. *Palaentology* **41**, 939–974 (1998).
16. Cherns, L. *Chelodes* and closely related Polyplacophora (Mollusca) from the Silurian of Gotland, Sweden. *Palaentology* **41**, 545–573 (1998).
17. Wingstrand, K. G. On the anatomy and relationships of Recent Monoplacophora. *Galathea Rep.* **16**, 7–94 (1985).

Supplementary information in the form of computer video files is available on *Nature's* World-Wide Web site (<http://www.nature.com>).

Acknowledgements

This work was supported by NERC. We acknowledge technical assistance by J. Hay, and we thank L. Cherns, J. Peel and J. Taylor for discussions.

Correspondence and requests for materials should be addressed to Derek J. Siveter (e-mail: derek.siveter@earth.ox.ac.uk).

Intraspecific competition favours niche width expansion in *Drosophila melanogaster*

Daniel I. Bolnick

Center for Population Biology, Storer Hall, University of California, Davis, California 95616, USA

Ecologists have proposed that when interspecific competition is reduced, competition within a species becomes a potent evolutionary force leading to rapid diversification¹. This view reflects the observation that populations invading species-poor communities frequently evolve broader niches². Niche expansion can be associated with an increase in phenotypic variance^{3,4} (known as character release⁵), with the evolution of polymorphisms^{6–9}, or with divergence into many species using distinct resources^{10–13} (adaptive radiation). The relationship between intraspecific competition

and diversification is known from theory^{14,15}, and has been used as the foundation for some models of speciation^{16–20}. However, there has been little empirical proof that niches evolve in response to intraspecific competition. To test this hypothesis, I introduced cadmium-intolerant *Drosophila melanogaster* populations to environments containing both cadmium-free and cadmium-laced resources. Here I show that populations experiencing high competition adapted to cadmium more rapidly than low competition populations. This provides experimental confirmation that competition in a population can drive niche expansion onto new resources for which competition is less severe.

I maintained ten cages of *D. melanogaster* populations, each cage containing resources varying in quality as a result of cadmium chloride added in varying concentrations to standard cornmeal fly medium. Each cage contained a roughly normal distribution of cadmium concentrations, obtained by varying the number of 10 ml vials of a given concentration that were placed in the cage (Fig. 1a). This design replicates the conditions of a model of competitive speciation¹⁷, in which a population is initially adapted to one tail of a resource distribution.

When a population is adapted to use only a subset of the resources available in its environment, frequency-dependent competition may favour niche expansion¹⁵. Frequency-dependent competition occurs when different phenotypes in a population use distinct subsets of the available resources, owing to functional trade-offs in using alternate resources. I chose to use cadmium because flies adapted to high cadmium have depressed fecundity and survivorship on unpolluted medium²¹, generating the requisite trade-offs. Competition is then a function of the density of similar phenotypes, rather than the total population size, although increasing and decreasing total population size will still change the intensity of frequency-dependent competition. In this experiment, competition is frequency dependent in the sense that the number of adults or larvae competing in a given food vial depends on the frequency of flies that can tolerate the vial's cadmium content. If fitness depression from competition in low cadmium vials exceeds the cost of cadmium exposure, then selection should favour adaptation to the new resource¹⁴. The selection differential between cadmium-tolerant and cadmium-intolerant phenotypes depends on the intensity of competition. Hence, populations experiencing higher competition (HC) should evolve more rapidly than low competition populations (LC) to tolerate cadmium.

Starting populations had severely reduced performance on high cadmium medium. No-choice assays revealed a 94.6% reduction in the number of progeny emerging from 50 $\mu\text{g ml}^{-1}$ cadmium, and a 99.4% reduction from 100 $\mu\text{g ml}^{-1}$, relative to 0 $\mu\text{g ml}^{-1}$ vials (Fig. 1a). At the start of this experiment, 0 $\mu\text{g ml}^{-1}$ cadmium vials were in disproportionately high demand because cadmium-intolerant flies dominated the population. The rare cadmium-tolerant individuals experienced reduced competition because they had access to abundant exclusive resources. Consequently, fly lines evolved a greater resistance to cadmium as indicated by the increasing number of flies emerging from high cadmium vials in the second and fourth generations (Fig. 1b, c). By the fourth generation, HC populations had significantly higher emergence rates on high cadmium than LC populations (Fig. 1c). The populations developed elevated performance at an unexpected speed. In understanding this remarkable rate of evolution, it is useful to note that in generation two, a mean of 154.2 flies emerged from 0 $\mu\text{g ml}^{-1}$ cadmium vials in each HC cage. With about 250 females per cage (assuming equal sex ratios), this yields a mean of 0.61 progeny per female if all flies avoided toxic medium, a fecundity an order of magnitude lower than that of the initial populations on 50 $\mu\text{g ml}^{-1}$ cadmium. This differential suggests that intense selection favoured cadmium-tolerant individuals. Some standing genetic variation for cadmium tolerance must have been present in the starting populations for such a

rapid response to selection.

Two possible models could explain the observed increase in fly emergence on cadmium. (1) A purely behavioural increase in oviposition on high cadmium in HC could result in higher emergence even if larval survivorship remains unchanged. Choosing to oviposit on cadmium reduced adult mortality, as many flies were trampled to death on the 0 $\mu\text{g ml}^{-1}$ medium in the first two generations. (2) The elevated eclosion rate on cadmium might reflect an increased larval physiological tolerance of cadmium, indicating that the populations actually evolved to use new resources more effectively. Two lines of evidence support the latter conclusion. First, HC flies were significantly heavier than LC flies when they emerged from 100 $\mu\text{g ml}^{-1}$ cadmium medium (two-tailed *t*-test: *t* = 3.223, 8 degrees of freedom (d.f.), *P* = 0.012).

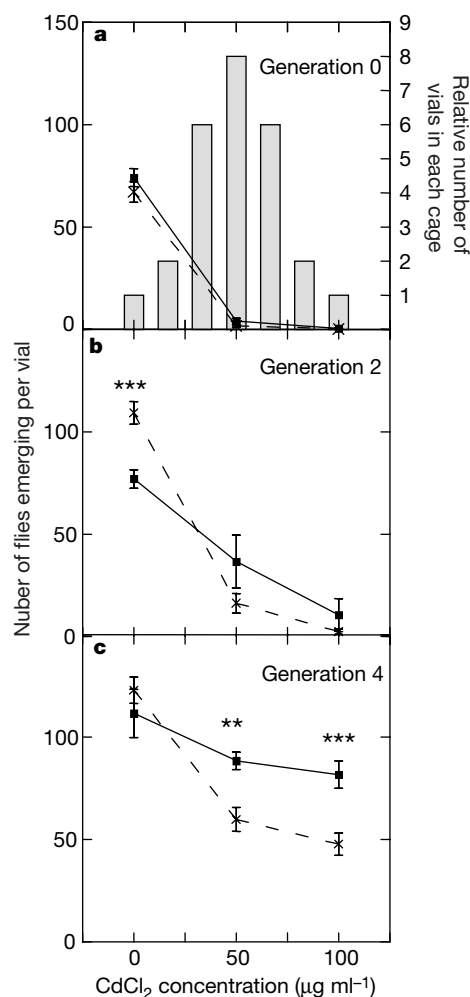


Figure 1 High competition *Drosophila* populations (squares) adapt to cadmium more rapidly than low competition populations (crosses). **a**, Starting lines had low fecundity on cadmium (ANOVA cadmium effect $F_{2,16} = 318$, $P = 0.0001$). The histogram shows the relative availability of each cadmium concentration within a cage. Note that starting populations performed poorly on most resources in each cage. **b**, In the second generation, eclosion rates in HC cages are significantly depressed on 0 $\mu\text{g ml}^{-1}$ cadmium because of larval and adult crowding (cadmium $F_{2,16} = 94.7$, $P < 0.0001$; cadmium $\times$ competition $F_{2,16} = 8.7$, $P = 0.003$). **c**, By generation 4, HC lines are significantly better adapted to cadmium, alleviating crowding on low cadmium (cadmium $F_{2,16} = 25.8$, $P < 0.0001$; cadmium $\times$ competition $F_{2,16} = 3.9$, $P = 0.042$; cadmium $\times$ cage (competition) $F_{16,30} = 3.36$, $P = 0.002$). The interaction between cadmium and cage nested within competition reflects variation among HC cages, which had uniformly higher emergence rates than LC cages. Tukey test comparisons: two asterisks, $P \leq 0.01$; three asterisks, $P \leq 0.001$. Bars indicate s.e.m.

Increased fly mass at eclosion correlates with the evolution of cadmium tolerance in some studies²¹. Second, a cross-rearing experiment showed that HC lines have heritable elevated resistance to cadmium.

In the cross-rearing experiment, I mated flies from high or low cadmium, and placed their eggs on high or low cadmium to test egg-to-adult survivorship. Mated pairs that emerged from $100 \mu\text{g ml}^{-1}$ in HC cages produced eggs with elevated survivorship on cadmium (Fig. 2). Progeny of HC flies from $0 \mu\text{g ml}^{-1}$ cadmium were not significantly different from progeny of LC flies (from 0 or $100 \mu\text{g ml}^{-1}$). Hence, there is a genetic distinction between flies emerging from different media within HC cages (separate-variance $t = 4.241$, 5.9 d.f., $P = 0.006$). For this to occur, there must either be maternal effects in which cadmium-raised flies have cadmium-tolerant progeny, or a link between cadmium tolerance and oviposition behaviour. Elevated HC survivorship on cadmium is not a result of maternal effects, as LC flies emerging from the same high concentration do not show elevated survivorship. Competition-induced maternal effects on survivorship are ruled out by the fact that competition treatment does not have an effect on survivorship of progeny reared on $0 \mu\text{g ml}^{-1}$ (separate-variance $t = 0.417$, 5.5 d.f., $P = 0.693$).

Choice trials on generation 4 flies showed no significant effect of larval environment (competition or cadmium) on oviposition preference, contrary to the hypothesis of linkage between tolerance and oviposition behaviour. However, the choice trials took place at low fly density, and so may not be a good gauge of fly behaviour in the population cages if resistant flies switch to cadmium only when

preferred toxin-free medium is crowded. Oviposition/resistance linkage was not expected, but is likely to have facilitated niche expansion in this experiment. The diversifying effect of frequency-dependent selection is exaggerated by decreased niche overlaps between phenotypes, as in when distinct resistance phenotypes lay eggs in different vials.

Cadmium resistance did not come at a cost of reduced survivorship on low cadmium. Although flies emerging from $100 \mu\text{g ml}^{-1}$ medium in HC populations had higher cadmium tolerance, they did not perform significantly worse on $0 \mu\text{g ml}^{-1}$ medium (compared with $0 \mu\text{g ml}^{-1}$ LC lines: $t = 0.837$, 6.8 d.f., $P = 0.432$; compared with $0 \mu\text{g ml}^{-1}$ HC lines: $t = 0.321$, 6.7 d.f., $P = 0.758$). This is consistent with results showing that cadmium adaptation lowered fecundity and fly mass, but not survivorship²¹.

High-density populations of *D. melanogaster* rapidly evolved to use cadmium-laced medium, as shown by higher emergence rates, higher fly mass, and heritable elevated survivorship. This result confirms existing theory¹⁴ that intraspecific competition can generate selection for niche expansion and associated phenotypic diversification. In particular, the experiment qualitatively matched the dynamics of a model of competitive speciation¹⁷, in which a population using only the tail of a resource distribution evolved to use the modal resource.

Previous studies have shown that frequency-dependent competition is important in maintaining stable polymorphisms in both experimental²² and natural populations^{23,24}. My results complement such studies by showing that competition also has a role in favouring the initial evolution of greater phenotypic variance, and in the rate of niche expansion. Competition favours diversification by depressing the fitness of individuals using the ancestral resource to the point where previously suboptimal resources have a higher value. Cadmium imposes decreased fecundity, survivorship, and fly mass even on 'adapted' flies. Cadmium's toxicity can be viewed as equivalent to the fitness costs of using any resource that a consumer is not adapted to use efficiently. The apple maggot fly, *Rhagoletis pomonella*, has reduced fecundity on apples compared with its ancestral resource, the hawthorn fruit²⁵. However, competition is more intense in hawthorn fruits, and may have driven some populations to specialize on apples²⁶. Many taxa adapt to new resources that are defended by toxins, such as *Drosophila sechellia*, which has adapted to the toxic fruits of *Morinda citrifolia*²⁷. The results reported here, combined with retrospective studies in natural populations, suggest that competition can favour adaptation to fitness valleys. Frequency-dependent effects argue for a more fluid vision of an 'adaptive landscape'²⁸, in which fitness peaks can be depressed by competition, to be level with or below what once were fitness valleys constraining phenotypic evolution²⁹. A fluid adaptive landscape appears far more likely to produce population diversification and adaptive radiation. □

Methods

Fly lines

Ten populations were constructed using twenty isogenic lines derived from a wild population (Univ. Orchard, Winters CA, maintained for 25 generations by full-sib mating) provided by S. Nuzhdin. These lines were outcrossed with each other for 2 generations and then mixed to produce 10 cages with genetically similar starting populations of 500 flies. Choice trials showed that flies avoid cadmium exposure, and can distinguish between concentrations rather than just presence/absence of the toxin. I put 1 vial of $16 \mu\text{g ml}^{-1}$ and 1 vial of $83 \mu\text{g ml}^{-1}$ cadmium medium in each of 36 observation chambers with 10 flies. After 2 d, counts of flies ovipositing or feeding on each medium type showed a strong preference for the lower cadmium food (an average of 1.11 versus 0.16 flies per vial, $\chi^2 = 15.37$, 1 d.f., $P < 0.001$).

Experimental design

I generated a resource distribution by varying the ratio of vials of different cadmium concentrations (7 levels: 0, 16, 33, 50, 66, 83 and $100 \mu\text{g ml}^{-1}$ in a 1:2:6:8:6:2:1 ratio within each cage, Fig. 1a). Vials were uncovered within a cage so flies could oviposit on any medium. To vary the level of competition, cages received a total of 26 or 104 vials of medium, resulting in an average of 9.6 or 2.4 females, respectively, per vial (high and low

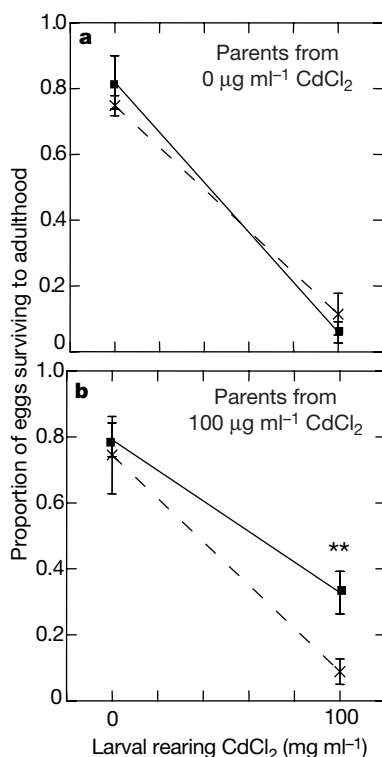


Figure 2 Survivorship of eggs from matings of parents emerging from different cadmium concentrations. **a**, $0 \mu\text{g ml}^{-1}$ cadmium; **b**, $100 \mu\text{g ml}^{-1}$ cadmium. Progeny of flies from $100 \mu\text{g ml}^{-1}$ cadmium in HC populations (squares) have higher survivorship on cadmium than progeny of flies from $100 \mu\text{g ml}^{-1}$ in LC populations (crosses) (separate variance $t = 3.626$, 6.5 d.f., $P = 0.01$), showing that HC lines evolved greater heritable cadmium tolerance. The survivorship of progeny of flies taken from $0 \mu\text{g ml}^{-1}$ cadmium sources (**a**) is identical between competition treatments, and identical to progeny of parents from LC $100 \mu\text{g ml}^{-1}$ (**b**). Bars indicate s.e.m. of means for each population. Two asterisks, $P \leq 0.01$.

competition, HC and LC; 5 cages per treatment). This design allows a constant census population size across treatments. Effective population size could be significantly smaller in HC cages because of elevated selection intensity. Each generation, parents mated for 5 d in a cage before being removed. Vials were left uncovered so emerging progeny could mix freely in their cage, and a random sample of 500 progeny from a given cage became the next generation's parents, with fresh arrays of medium. Cadmium slows development in unadapted flies²¹, so new parents were not collected until larvae were eclosing from even the highest cadmium vials, at which time many were still emerging from 0 cadmium as well. This scheme should minimize selection on development time and the influence of inbreeding among the flies that eclosed earliest.

Performance assays

To test performance of the starting lines (generation 0), subsamples of 6 adults from each population were mated in vials of 0, 50 and 100 $\mu\text{g ml}^{-1}$ medium (5 vials $\times$ 3 concentrations $\times$ 10 populations), and the number of eclosing progeny counted to estimate concentration-specific fecundity. In contrast to this no-choice design, assays of later generations confound the effects of female oviposition behaviour and larval survival. Each generation after removing the ovipositing parents, I stoppered two vials of 0, 50 and 100 $\mu\text{g ml}^{-1}$ medium within each cage, while all other vials were left open for emerging adults to mix in the cage. On removing adults to replace the medium tray and begin the following generation, I counted progeny caught within each stoppered vial. I also weighed 200 flies from each stoppered vial to calculate mean fly mass. In HC cages where only 1 vial of 0 and 100 $\mu\text{g ml}^{-1}$ was present, an extra vial of each was set in the cage for assay purposes, but those progeny were not returned to the population. Data in each generation were analysed with a nested mixed model (cage nested within competition crossed factorially with cadmium) analysis of variance (ANOVA). Residuals matched assumptions of the ANOVA.

Cross-rearing experiment

From each cage, I collected 5 fourth-generation male/female pairs from stoppered 0 $\mu\text{g ml}^{-1}$ cadmium vials and 5 from 100 $\mu\text{g ml}^{-1}$ cadmium vials. Each pair mated on grape agar, from which I collected 20 eggs per pair, placing half in one 0 $\mu\text{g ml}^{-1}$ vial and half in one 100 $\mu\text{g ml}^{-1}$ cadmium vial to produce a fully factorial cross-transplant design (10 populations $\times$ 2 source concentrations $\times$ 5 pairs $\times$ 2 rearing concentrations). The number of eclosing adults relative to the initial number of eggs measures survivorship. Hypothesis tests treated populations rather than parent pairs as the level of replication, taking the mean across all five parent pairs in a given cage $\times$ source $\times$ rearing combination. This approach is more robust than an ANOVA for this experimental design^{21,30}, as variances are unequal between cells. Accordingly each source $\times$ rearing medium combination has five HC population and five LC population replicates, which I compared with a separate-variance two-group *t*-test. A hierarchical ANOVA yields qualitatively similar results but with less conservative *P* values (for example, source medium $\times$ rearing medium $\times$ competition treatment $F_{1,166} = 7.731$, $P = 0.006$).

Received 21 August 2000; accepted 4 January 2001.

1. Simpson, G. G. *The Major Features of Evolution* (Columbia Univ. Press, New York, 1954).
2. Werner, T. K. & Sherry, T. W. Behavioral feeding specialization in *Pinaroloxia inornata*, the "Darwin's Finch" of Cocos Island, Costa Rica. *Proc. Natl Acad. Sci. USA* **84**, 5506–5510 (1986).
3. Van Valen, L. Morphological variation and width of ecological niche. *Am. Nat.* **49**, 377–389 (1965).
4. Lister, B. C. The nature of niche expansion in West Indian *Anolis* lizards. I: ecological consequences of reduced competition. *Evolution* **30**, 659–676 (1976).
5. Grant, P. R. Convergent and divergent character displacement. *Biol. J. Linn. Soc.* **4**, 39–69 (1972).
6. Robinson, B. W. & Schluter, D. in *Adaptive Genetic Variation in the Wild* (eds Mousseau, T., Sinervo, B. & Endler, J. A.) 65–94 (Oxford Univ. Press, New York, 2000).
7. Gislason, D., Ferguson, M. M., Skúlason, S. & Snorrason, S. S. Rapid and coupled phenotypic and genetic divergence in Icelandic arctic char (*Salvelinus alpinus*). *Can. J. Fish. Aquat. Sci.* **56**, 2229–2234 (1999).
8. Knerer, G. & Atwood, C. E. Diprionid sawflies: polymorphism and speciation. *Science* **179**, 1090–1099 (1973).
9. Meyer, A. Ecological and evolutionary consequences of the trophic polymorphism in *Cichlasoma citrinellum* (Pisces: Cichlidae). *Biol. J. Linn. Soc.* **39**, 279–299 (1990).
10. Fryer, G. & Iles, T. D. *The Cichlid Fishes of the Great Lakes of Africa* (Oliver and Boyd, London, 1972).
11. Kambysellis, M. P. & Craddock, E. M. in *Molecular Evolution and Adaptive Radiation* (eds Givnish, T. J. & Sytsma, K. J.) 475–509 (Cambridge Univ. Press, New York, 1997).
12. Lack, D. *Darwin's Finches* (Cambridge Univ. Press, Cambridge, 1947).
13. Schliewen, U. K., Tautz, D. & Paabo, S. Sympatric speciation suggested by monophyly of crater lake cichlids. *Nature* **368**, 629–632 (1994).
14. Wilson, D. S. & Turelli, M. Stable underdominance and the evolutionary invasion of empty niches. *Am. Nat.* **127**, 835–850 (1986).
15. Roughgarden, J. Evolution of niche width. *Am. Nat.* **106**, 683–718 (1972).
16. Rosenzweig, M. L. Competitive Speciation. *Biol. J. Linn. Soc.* **10**, 275–289 (1978).
17. Dieckmann, U. & Doebeli, M. On the origin of species by sympatric speciation. *Nature* **400**, 354–357 (1999).
18. Doebeli, M. A quantitative genetic competition model for sympatric speciation. *J. Evol. Biol.* **9**, 893–909 (1996).
19. Gibbons, J. R. H. A model for sympatric speciation in *Megarhyssa* (Hymenoptera: Ichneumonidae): competitive speciation. *Am. Nat.* **114**, 719–741 (1979).
20. Drosel, B. & McKane, A. Competitive speciation in quantitative genetic models. *J. Theor. Biol.* **204**, 467–478 (2000).
21. Shirley, M. D. F. & Sibly, R. M. Genetic basis of a between-environment trade-off involving resistance to cadmium in *Drosophila melanogaster*. *Evolution* **53**, 826–836 (1999).

22. Rainey, P. B. & Travisano, M. Adaptive radiation in a heterogeneous environment. *Nature* **394**, 69–72 (1998).
23. Hori, M. Frequency-dependent natural selection in the handedness of scale-eating cichlid fish. *Science* **260**, 216–219 (1993).
24. Benkman, C. W. Are the ratios of bill crossing morphs in crossbills the result of frequency-dependent selection? *Evol. Biol.* **10**, 119–126 (1996).
25. Prokopy, R. J., Diehl, S. R. & Cooley, S. S. Behavioral evidence for host races in *Rhagoletis pomonella*. *Oecologia* **76**, 138–147 (1988).
26. Feder, J. L., Reynolds, K., Go, W. & Wang, E. C. Intra- and interspecific competition and host race formation in the apple maggot fly, *Rhagoletis pomonella* (Diptera: Tephritidae). *Oecologia* **101**, 416–425 (1995).
27. Jones, C. D. The genetic basis of *Drosophila sechellia*'s resistance to a host plant toxin. *Genetics* **149**, 1899–1908. (1998).
28. Fisher, R. A. *The Genetical Theory of Natural Selection* (Dover, New York, 1958).
29. Coyne, J. A., Barton, N. H. & Turelli, M. Perspective: A critique of Sewall Wright's shifting balance theory of evolution. *Evolution* **51**, 643–671 (1997).
30. Leroi, A. M., Chippindale, A. K. & Rose, M. R. Long-term laboratory evolution of a genetic life-history trade-off in *Drosophila melanogaster*. 1. The role of genotype-by-environment interaction. *Evolution* **48**, 1244–1257 (1994).

Acknowledgements

I thank S. Nuzhdin, M. Doebeli, R. Haygood, M. Servedio and B. Holland for discussions on experimental design. P. Wescott, E. Barker and E. Baumgartner provided invaluable laboratory help. This research was supported by a Daphne and Ted Pengelly Grant (UC Davis), an NSF grant to S. Nuzhdin, and an NSF Pre-doctoral Fellowship.

Correspondence and requests for materials should be addressed to D.I.B. (e-mail: dibolnick@ucdavis.edu).

Odour-plume dynamics influence the brain's olfactory code

Neil J. Vickers^{*†}, Thomas A. Christensen^{*}, Thomas C. Baker[‡] & John G. Hildebrand^{*}

^{*} Arizona Research Laboratories Division of Neurobiology, The University of Arizona, PO Box 210077, Tucson, Arizona 85721, USA

[‡] Department of Entomology, Iowa State University, Ames, Iowa 50011, USA

The neural computations used to represent olfactory information in the brain have long been investigated^{1–3}. Recent studies in the insect antennal lobe suggest that precise temporal and/or spatial patterns of activity underlie the recognition and discrimination of different odours^{3–7}, and that these patterns may be strengthened by associative learning^{8,9}. It remains unknown, however, whether these activity patterns persist when odour intensity varies rapidly and unpredictably, as often occurs in nature^{10,11}. Here we show that with naturally intermittent odour stimulation, spike patterns recorded from moth antennal-lobe output neurons varied predictably with the fine-scale temporal dynamics and intensity of the odour. These data support the hypothesis that olfactory circuits compensate for contextual variations in the stimulus pattern with high temporal precision. The timing of output neuron activity is constantly modulated to reflect ongoing changes in stimulus intensity and dynamics that occur on a millisecond timescale.

In order to examine the relationship between the fine-scale structure of a natural odour plume and the complex patterns of odour-evoked activity in the brain, we used a well-characterized olfactory subsystem in male moths that is devoted to processing information about the female sex pheromone^{11–14}. First, an electro-antennogram (EAG) was used to monitor the spatiotemporal properties of a pheromone plume generated inside a laboratory wind tunnel (Fig. 1a). Surprisingly, the structure of the odour

[†] Present address: Department of Biology, University of Utah, Salt Lake City, Utah 84112, USA.

plume varied substantially at different sampling points separated by as little as 5 cm (Fig. 1b). At low wind speeds, both the largest EAG bursts and the most frequent fluctuations in EAG activity occurred in the central zone of the plume (Fig. 1c, centre), with virtually no activity at the peripheral sampling sites. As wind speed increased, however, the plume became more dispersed, with larger-amplitude EAG bursts occurring more frequently at the peripheral sites. Burst amplitude generally decreased as wind speed increased, and the location of the largest-amplitude responses shifted from the plume centre to the periphery (Fig. 1c). At the highest wind speed, the centre of the plume was characterized by higher-frequency, lower-amplitude bursts (Fig. 1c). Thus, in a natural airborne plume, odour contacts with the antennae are brief and frequent (up to 5 s^{-1}), and antennal responses furthermore show significant variation in amplitude over the course of stimulation (Fig. 1b and c). These intrinsic plume dynamics mean that very minor shifts in the position of a stationary antenna relative to the odour source resulted in large differences in the detected spatiotemporal pattern of odour striking the antenna.

To address the issue of how these spatiotemporal patterns might be further modified when an animal engages in typical locomotory activity (such as flight), male moths (*Heliothis virescens*) were fitted with a third antenna as an EAG monitor, and then allowed to fly upwind in a natural pheromone plume¹⁵. The EAG activity thus recorded was then synchronized to a video recording of the moth's flight track (Fig. 2a, b). These moths were tested under conditions identical to those used to measure plume dynamics at 60 cm s^{-1} with a stationary antenna (Fig. 1b, c). Sampling across 20 individual flights revealed that males encountered odour filaments most frequently in the centre of the plume, with significantly fewer contacts recorded toward the plume edges (Fig. 2c, left). In contrast, the amplitude distribution of in-flight plume contacts was considerably more uniform across the plume (Fig. 2c, right). There was also a marked difference in the mean EAG pulse amplitudes evoked

at rest and in flight, the latter being 2–3 times greater than the former (compare Figs 1c and 2c). These results indicate that the animal's own in-flight movements influence the detection of odour pulses striking the antenna, and that the initial sensory representations of the odour stimulus in the brain may vary considerably depending on the animal's level of activity.

Next, we examined how these rapidly changing spatiotemporal patterns affect the representation of olfactory information in the glomerular circuits in the insect's brain. In male moths, the highly attractive sex-pheromone blend is readily discriminated from deterrent blends released by related species^{11–14}. This is accomplished through the selective activation of neurons associated with specific combinations of glomeruli in the male's macroglomerular complex (MGC)¹⁶. Recordings were made simultaneously from glomerular output neurons (projection neurons; PNs) and the ipsilateral antenna (EAG) in response to point-source odour plumes created inside a small wind tunnel (Fig. 3). Extended recordings from four PNs that responded selectively to only one specific component of the sex-pheromone blend revealed that glomerular output is tightly synchronized to the temporal pattern of peripheral input to the olfactory system (Fig. 3a). Analysis of the variations in instantaneous PN spike frequency as a function of EAG burst amplitude reveals both a progressive increase in the number of spikes evoked as stimulus intensity increases (Fig. 3c) as well as considerable temporal variation in the PN response from trial to trial (Fig. 3b). Closer examination of the temporal structure of PN responses reveals no reproducible pattern of spike timing over dozens of repeated odour presentations (Fig. 4a). Responses vary from regular, tonic firing patterns at low stimulus intensities to more complex phasic-tonic spike trains at higher intensities (Fig. 3a, b). The time interval between maximum EAG amplitude and maximum instantaneous PN frequency also varies as a function of stimulus intensity, from 58.6 ms at low intensities to 14.0 ms at the highest intensities (mean $\pm$ standard deviation,

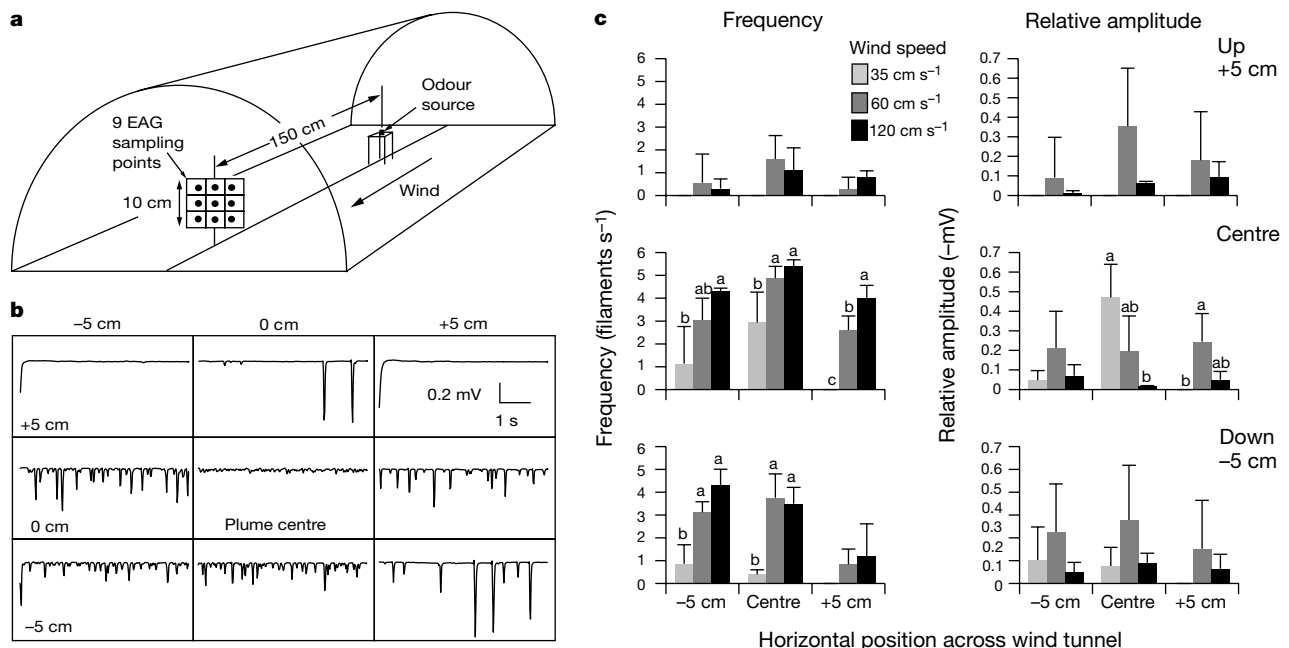


Figure 1 Mapping the dynamics of a natural odour plume. **a**, The electroantennogram (EAG) was used to monitor odour-evoked activity in response to a wind-borne odour plume generated from a point source in a wind tunnel. The positions of the odour source and the nine sequential EAG sampling points 150 cm downwind are indicated. Separation between sampling points was small (5 cm) so that the fine-scale spatiotemporal structure of the plume could be measured. **b**, EAG activity of a single antenna recorded at the nine sampling points, showing position-dependent variation in the amplitude and frequency of

depolarizations (bursts) in response to odour strands over a 5-s period at a wind speed of 60 cm s^{-1} . **c**, Summary of mean frequency and relative amplitude data illustrating the spatiotemporal dynamics of the odour-evoked EAG activity across the entire nine-point array at different wind speeds ($n = 3, 5$, and 2 for wind speeds of $35, 60$, and 120 cm s^{-1} , respectively). Means ($\pm$ s.d.) at each location were compared across wind speeds (MANOVA, followed by a post-hoc Newman–Keuls test, $P < 0.05$).

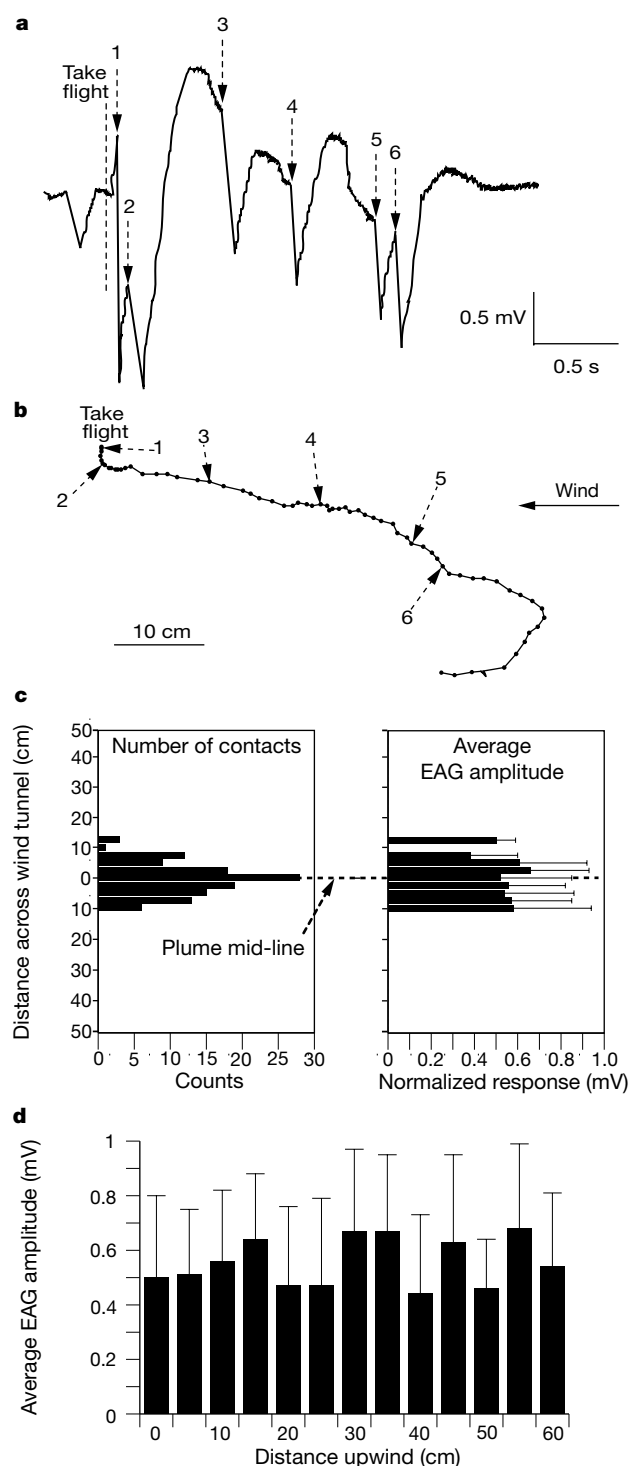


Figure 2 Measurement of plume dynamics in flight. **a, b**, Simultaneously recorded EAG and flight-track data from an *H. virescens* male. Onsets of EAG bursts are labelled with numbers that correspond to positions of the male along the flight track. Wind direction and the position of the plume midline are indicated by an arrow. We note that all EAG activity recorded during the flight of this male occurred within ± 6 cm of the plume midline. **c**, Plot of the positions across the wind tunnel (viewed from above) where males encountered pheromone strands, as indicated by bursts of EAG activity. The number of contacts (left) ($n = 20$ flights) was greatest along the plume midline and decreased rapidly to zero beyond approximately 10 cm to either side of the plume midline, even though flights extended over nearly the entire 1-m width of the wind tunnel. Averaged measurements of EAG relative amplitudes (mean $\pm$ s.d., right) revealed no significant variation (ANOVA, $P > 0.5$) across the plume. **d**, Normalized EAG amplitudes (mean $\pm$ s.d.), plotted against their positions along the long axis of the wind tunnel, showed no significant changes with distance from the odour source (ANOVA, $P > 0.5$).

s.d. = 35.6 ± 13.3 ms; $n = 44$). These results were observed whether the PN responded selectively to the major pheromone component ($n = 2$) or to the second essential component ($n = 2$) in the *H. virescens* blend¹⁶.

Different parameters of the PN response were also shown to be strongly correlated to different time-dependent stimulus features. First, spike-train onset in PNs is closely synchronized to EAG onset, occurring at a nearly constant latency of 44.3 ± 1.7 ms ($n = 165$), regardless of EAG amplitude. Second, both the mean rate and the maximum instantaneous rate of PN spiking are proportional to EAG burst amplitude (Fig. 3c, left), but neither of these parameters varies significantly with its rate of change (EAG onset slope; Fig. 3c, right). Thus, in the moth olfactory system, the time course of PN spike patterns is strongly dependent on both stimulus intensity and plume dynamics.

Synchronous firing among neurons is widely believed to facilitate the integration of related signals in a neural network, thus creating a stronger, more coherent representation of a given stimulus in the brain. In the olfactory system, recent evidence for odour-evoked oscillatory synchronization of PN ensembles has been obtained in several insect species^{3–5,9}. Oscillations of brain potentials are also a prominent component of olfactory network activity in moths⁷, and thus we wanted to know whether PNs in the moth antennal lobe are also synchronized by oscillations, or whether synchrony is governed by alternate mechanisms. As a first step, we conducted a temporal analysis of PN responses evoked by many consecutive odour pulses of varying duration and amplitude in a point-source plume (Fig. 4). If PNs were modulated by an evoked wave or network oscillation, PN spike activity would show a strong periodicity in the peristimulus histogram of evoked responses, as shown in other species^{3–5,9}. Instead, a histogram of PN instantaneous frequencies (Fig. 4b) revealed non-periodic activity distributed over a very wide frequency range (1.6–185.2 Hz), with little power in the slower frequency range of local field potential (LFP) oscillations (20–50 Hz) reported in moths⁷ and other insects^{3–5,9}. This finding was confirmed by Fourier analysis (Fig. 4c), which revealed a spectral maximum at only 2.2 Hz, reflecting not the LFP oscillation but the much slower and regular “burst return period” of odour filaments in the plume¹⁰. Separate studies in the moth *Manduca sexta* have demonstrated that coactivity among PNs is most prominent at response onset, and is neither modulated nor synchronized to any phase of the LFP oscillation (coactivity occurs even in response to brief stimulus pulses which preclude the development of oscillations)^{17,18} (see Supplementary Information).

Simultaneous intracellular and EAG recordings indicate, therefore, that the activity of at least some olfactory PNs is time-locked to stimulus dynamics (Figs 3 and 4). Furthermore, the temporal patterning of PN responses is not constrained by an LFP but is highly variable from stimulus to stimulus (Fig. 4). This patterning may be governed in part by an underlying oscillation, but our data indicate that PN response dynamics are more reflective of stimulus dynamics, including changes in stimulus time course and intensity.

The dynamic properties of olfactory stimuli dictate that odour-mediated behaviours must be executed under a wide range of stimulus conditions¹⁰. For some species, ongoing changes in the environment could have a significant impact on the detection of odour signals by peripheral receptors, and subsequently, on the neural integration of this information in the brain. Behavioural studies have revealed that a patchy olfactory stimulus is more effective than continuous stimulation in sustaining manoeuvres by flying insects to locate the odour source^{11–13,19}. In accordance with these studies, our new results provide additional evidence for context-dependent modulation of glomerular output at the very earliest stages of processing in the olfactory system¹⁷. These findings have implications for deciphering the neural computations used by the brain to discriminate different odours. In other insects, for

example, it was found that different odours activate specific PN assemblies in the antennal lobe, but the patterns of PN spikes are synchronized to a coherent network oscillation³. From these studies and others^{4,5,9}, it has been suggested that each odour may be uniquely represented by a different temporal pattern of activity across the PN array. Although these findings are intriguing, it remains to be clarified whether such temporal patterns in PN assemblies are maintained under conditions when the spatio-temporal pattern of the stimulus itself is changing rapidly and

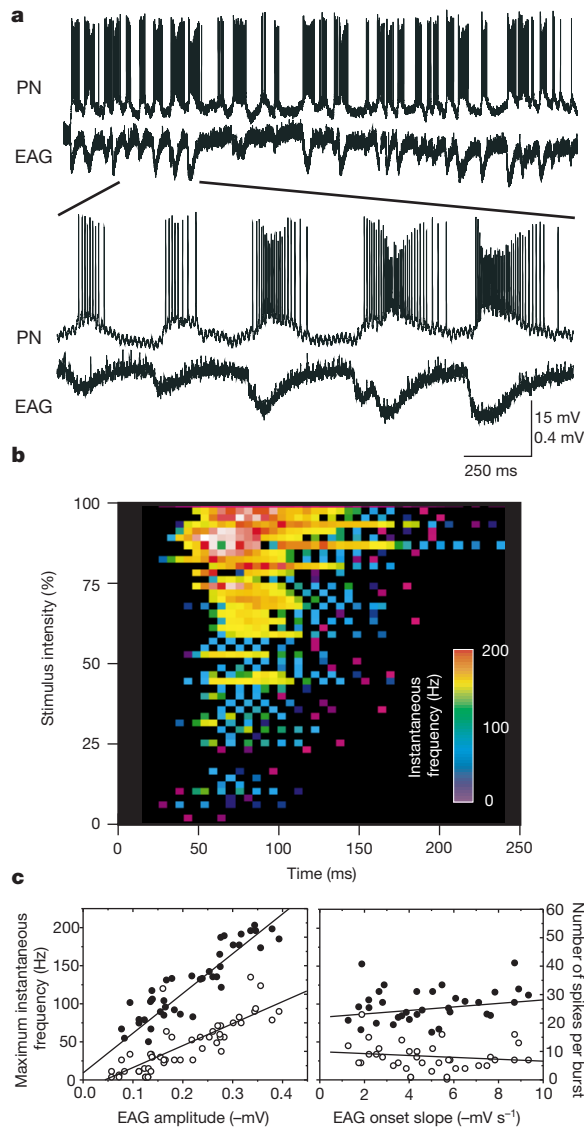


Figure 3 Temporal patterns of spike discharges in olfactory projection neurons (PNs) are strongly dependent on the stimulus dynamics of a wind-borne pheromone plume. **a**, A 15-s segment of activity recorded simultaneously from a PN and the ipsilateral antenna (EAG; upper pair of traces). Detail of the indicated segment is shown in the lower traces (timescale for lower PN-EAG traces only). Throughout the recording, PN activity is tightly correlated with antennal activation. **b**, Temporal variations in instantaneous spike frequency as a function of odour intensity (as represented by EAG burst amplitude). Raster plots from each response burst are colour-coded to reflect instantaneous spike frequency. We note that the response patterns become more temporally complex at elevated odour intensities, with increased number and frequency of spikes discharged. **c**, Intensity of odour stimulation (left) was tightly correlated to both the maximum instantaneous frequency of PN response (solid circles: $r^2 = 0.76$; $P < 0.0001$) and the number of spikes per odour-evoked burst (open circles: $r^2 = 0.65$; $P < 0.0001$). Other measures of activity such as EAG onset slope (right) did not serve as reliable predictors of central activity.

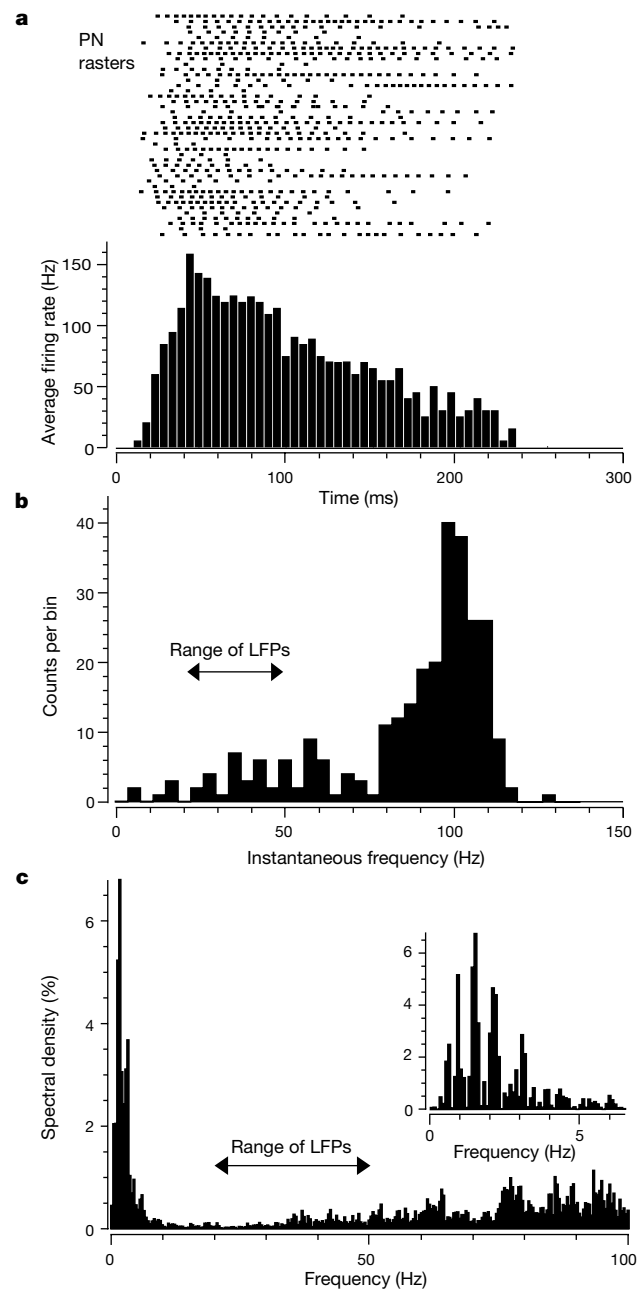


Figure 4 Temporal analysis of stimulus-evoked PN activity reveals that odour identity is not encoded in the temporal structure of these responses. **a**, Sequence of raster plots (upper) from a 20-s recording showing the consecutive responses to 40 odour pulses of varying intensity arising from a point source in the wind tunnel. Time course and number of spikes in each response (3 to 28) were strongly correlated with stimulus intensity, as in Fig. 3a. Summary histogram (lower) shows the average response over time (5-ms bins). Responses are characterized by a rapid rising phase that reaches a single peak in less than 50 ms from onset. This is followed by a slower recovery phase which varies in duration depending on stimulus intensity. The absence of periodicity in spike timing across stimulus intensities further indicates that PN spikes are not tightly synchronized to a slow, repetitive network event like a wave or oscillation. **b**, Instantaneous frequency histogram showing a normal distribution of spike intervals, with a prominent peak near 100 Hz. We note that the great majority of events fall well outside the frequency range of local field potential (LFP) oscillations reported in moths⁷ and other insects^{3-5,9}. **c**, Spectral analysis of the data revealed the strongest focus of activity below 10 Hz. An expanded plot (inset) shows the spectral peak at 2.2 Hz, reflecting the slow period of spike bursts that characterizes the intermittent response pattern evoked by the fluctuating odour dynamics. The absence of significant peaks in the LFP frequency range again indicates that PN spikes are not significantly entrained to periodic network events such as waves or oscillations.

unpredictably, as occurs in natural odour plumes^{10–17}. While our results may be representative of olfactory networks that must monitor the temporal properties of odour plumes, the data show that in the moth antennal lobe, the timing of spike activity across the PN ensemble is not constrained by a periodic network event like an induced wave or oscillation¹⁸. Instead, timing remains flexible, allowing PN spike trains to be modulated from moment to moment to reflect the changing pattern of olfactory input to the brain. Recent observations in female moths suggest further that these mechanisms are not unique to the male moth's olfactory system or to pheromone-information processing^{13,20}. Our new findings therefore support previous studies in both invertebrate and vertebrate sensory systems, arguing that much of the information carried by spikes in sensory systems is associated with stimulus timing²¹. Our new results in moth PNs also support a model of olfactory multitasking in the antennal lobe²². We propose that it is through a combination of coding mechanisms that these output neurons are able to report information about both the chemical identity of the stimulus (through their specific spatial association with one or more glomeruli), and the ecologically relevant information about naturally occurring and unpredictable odour dynamics. □

Methods

Animals

Heliothis virescens (Lepidoptera: Noctuidae) were reared under controlled light (14 h light: 10 h dark), temperature (25–28 °C), and relative humidity (40–60%). All rearing procedures and dissection methods have been described previously^{16,21,23}. The procedures used to prepare moths for in-flight EAG measurements have been described elsewhere¹⁵.

Flight tunnel

The semicircular wind tunnel used for the stationary and in-flight EAG experiments with *H. virescens* males measured approximately 1 × 1 m in cross-section and 2 m in length. As a stimulus, 10 µl of a 1 mg µl⁻¹ solution of the species-specific six-component pheromone blend was pipetted onto a rubber septum¹⁵. The septum was positioned at a height of 15 cm, and a distance of 1.5 m upwind of the stationary EAG preparation (or the take-off platform in the case of the in-flight EAG experiments; see below). Wind speed was varied between 35 cm s⁻¹ and 120 cm s⁻¹ for the stationary EAG experiment, and was held constant at 60 cm s⁻¹ for the in-flight EAG experiments.

Electrophysiology

Essentially the same EAG preparation was used for both the stationary and in-flight experiments. The exposed and sharpened tips of two Teflon-coated silver wires (100 µm in diameter) were chloridized and inserted into each end of an excised male antenna and then secured with a tiny amount of wax. One wire was attached to the ground, the other connected through a small preamplifier headstage to a Syntech (Hilversum, The Netherlands) d.c. amplifier. For in-flight recordings, the EAG wires were 1.5 m in length, permitting recordings across the entire wind tunnel. For stationary EAG recordings, the preparation was moved through a nine-point sampling grid (Fig. 1a) centred on the plume originating 1.5 m upwind. Each sampling position was monitored for approximately 10 s until the d.c. shift in the trace stabilized¹⁵. The data reported here were collected for 5 s after trace stabilization.

Paired EAG/intracellular recordings

Glass micropipettes were used to make recordings from the dendrites of MGC (macroglomerular complex) output neurons^{16,23}. Electrodes were filled with 4% Lucifer Yellow CH (Sigma Chemical Company) in 0.2 M LiCl (backfilled with 2 M LiCl) so that neurons could be morphologically identified. Once impaled, each neuron was tested with a standard set of odours and odour mixtures^{16,23}. After exposure of the brain and the antennal lobes, the preparation was positioned at the downwind end of another smaller (41 cm inner diameter) wind tunnel especially designed for these experiments. Just before paired EAG/intracellular recording, a rubber septum loaded with pheromone (as above) was placed into the wind tunnel at a distance of 70 cm upwind of the preparation. Wind speed was set at 35 cm s⁻¹ and the plume created by the point source was allowed to passively stimulate the combined EAG/intracellular recording preparation.

Analysis

Stationary EAG. Frequency and amplitude data were extracted from a 5-s segment of the EAG recordings at each sampling site in the nine-point array (Fig. 1a). Relative EAG amplitudes were calculated to allow quantitative comparison of responses across different antennae. Measurements were taken at three different windspeeds: 35, 60, and 120 cm s⁻¹.

In-flight EAG. Twenty males were used to obtain in-flight EAG recordings. Each flight track was carefully aligned with respect to pheromone plume position using several fixed markers on the wind tunnel floor. The positions along each track where males encountered a pheromone pulse (as indicated by the simultaneously recorded EAG traces) were marked. From these aligned flight tracks, a composite figure of the positions of EAG bursts for all 20 males was produced. The horizontal cross-section of the wind tunnel was divided into a grid: across-tunnel strips centred on the plume midline were 2.5 cm in width; along-tunnel strips were 5 cm in length. The number of EAG peaks occurring within each across-tunnel strip was then summed and plotted. Relative EAG amplitudes were also calculated and averaged for each across-tunnel (2.5-cm) strip and each along-tunnel (5-cm) strip.

Paired EAG/intracellular records. Simultaneous long-term intracellular and EAG recordings were obtained from four preparations, and both records were digitized using Axoscope software (Axon Instruments). Instantaneous frequencies of spike activity were calculated using an in-house algorithm (courtesy of J. K. Douglass). Various response parameters including numbers of spikes per burst, burst durations and intervals, EAG burst amplitudes and onset slopes were measured directly in Axoscope.

Received 7 August 2000; accepted 3 January 2001.

- Adrian, E. D. Olfactory reactions in the brain of the hedgehog. *J. Physiol.* **100**, 459–473 (1942).
- Gelperin, A. & Tank, D. W. Odour-modulated collective network oscillations of olfactory interneurons in a terrestrial mollusc. *Nature* **345**, 437–440 (1990).
- Laurent, G., Wehr, M. & Davidowitz, H. Temporal representations of odors in an olfactory network. *J. Neurosci.* **16**, 3837–3847 (1996).
- Laurent, G. & Davidowitz, H. Encoding of olfactory information with oscillating neural assemblies. *Science* **265**, 1872–1875 (1994).
- Wehr, M. & Laurent, G. Odour encoding by temporal sequences of firing in oscillating neural assemblies. *Nature* **384**, 162–166 (1996).
- Joerges, J., Küttner, A., Galizia, G. & Menzel, R. Representations of odours and odour mixtures visualized in the honeybee brain. *Nature* **387**, 285–288 (1997).
- Heinbockel, T., Kloppenburg, P. & Hildebrand, J. G. Pheromone-evoked potentials and oscillations in the antennal lobes of the sphinx moth *Manduca sexta*. *J. Comp. Physiol. A* **182**, 703–714 (1998).
- Faber, T., Joerges, J. & Menzel, R. Associative learning modifies neural representations of odours in the insect brain. *Nature Neurosci.* **2**, 74–78 (1999).
- Stopfer, M. & Laurent, G. Short-term memory in olfactory network dynamics. *Nature* **402**, 664–668 (1999).
- Murlis, J. in *Insect Pheromone Research: New Directions* (eds Cardé, R. T. & Minks, A. K.) 221–231 (Chapman & Hall, New York, 1997).
- Christensen, T. A., Heinbockel, T. & Hildebrand, J. G. Olfactory information processing in the brain: encoding chemical and temporal features of odors. *J. Neurobiol.* **30**, 82–91 (1996).
- Hildebrand, J. G. Olfactory control of behavior in moths: central processing of odor information and the functional significance of olfactory glomeruli. *J. Comp. Physiol. A* **178**, 5–19 (1996).
- Hansson, B. S. & Christensen, T. A. in *Insect Olfaction* (ed. Hansson, B. S.) 125–161 (Springer, Berlin, 1999).
- Christensen, T. A. & White, J. E. in *The Neurobiology of Taste and Smell* Vol. 2 (eds Finger, T. E., Silver, W. L. & Restrepo, D.) 201–232 (Wiley, New York, 2001).
- Vickers, N. J. & Baker, T. C. Reiterative responses to single strands of odor promote sustained upwind flight and odor source location by moths. *Proc. Natl Acad. Sci. USA* **91**, 5756–5760 (1994).
- Vickers, N. J., Christensen, T. A. & Hildebrand, J. G. Combinatorial odor discrimination in the brain: attractive and antagonist odor blends are represented in distinct combinations of uniquely identifiable glomeruli. *J. Comp. Neurol.* **400**, 35–56 (1998).
- Christensen, T. A., Pawlowski, V. M., Lei, H. & Hildebrand, J. G. Multi-unit recordings reveal context-dependent modulation of synchrony in odor-specific neural ensembles. *Nature Neurosci.* **3**, 927–931 (2000).
- Lei, H., Christensen, T. A. & Hildebrand, J. G. Odor-evoked synchronization of olfactory networks: comparison of output neurons innervating the same and different glomeruli in the antennal lobe of *Manduca sexta*. *Soc. Neurosci. Abstr.* **26**, 1208 (2000).
- Baker, T. C. & Vickers, N. J. in *Insect Pheromone Research: New Directions* (eds Cardé, R. T. & Minks, A. K.) 248–264 (Chapman & Hall, New York, 1997).
- King, J. R., Christensen, T. A. & Hildebrand, J. G. Response characteristics of an identified, sexually dimorphic olfactory glomerulus. *J. Neurosci.* **20**, 2391–2399 (2000).
- Buračas, G. T. & Albright, T. D. Gauging sensory representations in the brain. *Trends Neurosci.* **22**, 303–309 (1999).
- Christensen, T. A., Waldrop, B. R. & Hildebrand, J. G. Multitasking in the olfactory system: context dependent responses to odors reveal dual GABA-regulated coding mechanisms in single olfactory projection neurons. *J. Neurosci.* **18**, 5999–6008 (1998).
- Christensen, T. A., Mustaparta, H. & Hildebrand, J. G. Chemical communication in heliothine moths. VI. Parallel pathways for information processing in the macroglomerular complex of the male tobacco budworm moth *Heliothis virescens*. *J. Comp. Physiol. A* **177**, 545–557 (1995).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank S. Hannaford and M. Willis for help in constructing the small wind tunnel. This work was supported by United States Department of Agriculture (National Research Initiative), National Institutes of Health (National Institute for Deafness and Other Communication Disorders), and Defense Advanced Research Projects Agency (Controlled Biological Systems).

Correspondence and requests for materials should be addressed to N.J.V. (e-mail: vickers@biology.utah.edu).

Interleukin-1 β -mediated induction of Cox-2 in the CNS contributes to inflammatory pain hypersensitivity

Tarek A. Samad*, Kimberly A. Moore*, Adam Sapirstein†, Sara Billet*, Andrew Allchorne‡, Stephen Poole§, Joseph V. Bonventre|| & Clifford J. Woolf*‡

*Neural Plasticity Research Group, Department of Anesthesia & Critical Care,

†Department of Anesthesia & Critical Care and ||Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Charlestown, Massachusetts 02129, USA

‡Department of Anatomy, University College London, London WC1E 6BT, UK

§National Institute for Biological Standards and Control, Blanche Lane, South Mimms, Hertfordshire EN6 3QG, UK

Inflammation causes the induction of cyclooxygenase-2 (Cox-2)¹, leading to the release of prostanoids, which sensitize peripheral nociceptor terminals and produce localized pain hypersensitivity². Peripheral inflammation also generates pain hypersensitivity in neighbouring uninjured tissue (secondary hyperalgesia), because of increased neuronal excitability in the spinal cord (central sensitization)³, and a syndrome comprising diffuse muscle and joint pain, fever, lethargy and anorexia⁴. Here we show that Cox-2 may be involved in these central nervous system (CNS) responses, by finding a widespread induction of Cox-2 expression in spinal cord neurons and in other regions of the CNS, elevating prostaglandin E₂ (PGE₂) levels in the cerebrospinal fluid. The major inducer of central Cox-2 upregulation is interleukin-1 β in the CNS, and as basal phospholipase A₂ activity in the CNS does not change with peripheral inflammation, Cox-2 levels must regulate central prostanoid production. Intraspinal administration of an interleukin-converting enzyme or Cox-2 inhibitor decreases inflammation-induced central PGE₂ levels and mechanical hyperalgesia. Thus, preventing central prostanoid production by inhibiting the interleukin-1 β -mediated induction of Cox-2 in neurons or by inhibiting central Cox-2 activity reduces centrally generated inflammatory pain hypersensitivity.

The major therapy for inflammatory pain is non-steroidal anti-inflammatory drugs (NSAIDs), which reduce prostanoid levels⁵. Prostanoid production requires both free arachidonic acid substrate and Cox. The two rate-limiting steps in prostaglandin synthesis are the regulation of arachidonic acid release from membrane phospholipids by phospholipase A₂ (PLA₂) and the conversion of arachidonic acid to the prostanoid precursor PGH₂ by Cox⁵. Two isoforms of Cox have been identified; Cox-1 is constitutively expressed in many cell types, and Cox-2 is induced at the site of inflammation, although it is also constitutively expressed in the kidney and parts of the CNS¹. The structure of the Cox-2 gene promoter, containing a TATA box and several transcription factor binding sites, is consistent with the finding that Cox-2 behaves like an immediate-early response gene in several tissues, and that transcription is a key mechanism for its regulation⁶. Cox is the principal target for NSAIDs, and their anti-inflammatory and analgesic benefits have been proposed to derive from inhibition of only the Cox-2 isoform induced at the site of inflammation^{1,5}. Prostanoids contribute to the development of peripheral sensitization through protein kinase A-mediated phosphorylation of sodium channels in nociceptor terminals, increasing excitability and reducing the pain threshold². The analgesic action of NSAIDs may, however, include a central site^{7,8}, and prostanoids may have a hyperalgesic action in the spinal cord⁹. We have investigated how and where Cox-2 is regulated in the CNS, whether it is the

prime determinant of central prostanoid production, and whether prevention of its induction or action alters peripheral inflammation-induced pain hypersensitivity.

We first investigated the extent of Cox-2 induction in the CNS following unilateral hindpaw inflammation evoked by complete Freund's adjuvant (CFA). Local peripheral inflammation produces a time-dependent increase in Cox-2 messenger RNA in the region of the lumbar spinal cord where sciatic afferents terminate (L4/L5 segments; Fig. 1a). Ipsilateral induction is maximal at 6 h (16-fold) and persists for at least 24 h (9-fold), as reported^{10,11}. At early time points (2–6 h) the increase is greatest ipsilateral to the inflammation, but we find that by 12 and 24 h the levels are bilaterally symmetrical in the lumbar spinal cord (Fig. 1a), suggesting a generalized increase in Cox-2 expression. To test this, we examined Cox-2 mRNA expression in the CNS. Hindpaw inflammation produces a large increase in Cox-2 mRNA in the skin (18-fold), and a similar relative increase in the lumbar and cervical regions of the spinal cord (Fig. 1b). Significant increases were also observed in the pons, ventral midbrain and hypothalamus (data not shown), and a large maintained increase (12-fold) was observed in the thalamus (Fig. 1b). Constitutive Cox-2 expression in samples from the entire cerebral cortex was not significantly affected by inflammation (Fig. 1b), but this does not exclude changes in areas limited to somatosensory processing. Low basal levels of Cox-2 protein are detectable immunohistochemically in dorsal horn neurons, with substantial upregulation 6 and 12 h after inflammation in neuronal cells (Fig. 1c, d). Changes in spinal Cox-2 are also observed by western blot analysis (Fig. 1e). The induction of central Cox-2 expression is associated with a substantial (>80-fold) increase in PGE₂ in the cerebrospinal fluid, peaking 12 h after inflammation (Fig. 1f). There was a smaller, non-significant increase in PGD₂ in the cerebrospinal fluid (CSF; 200 $\pm$ 135 pg ml⁻¹ 12 h after inflammation from naive levels of 90.3 $\pm$ 37 pg ml⁻¹, $P > 0.05$, $n = 5$), illustrating that central prostanoid induction is selective.

These data may indicate that Cox-2 is a major contributor to the elevation of spinal PGE₂ after peripheral inflammation. Cox-1 is not induced in the spinal cord¹² and its inhibition does not reduce pain behaviour⁸. It is possible, however, that induction or activation of central PLA₂ also contributes to increased eicosanoid generation. The PLA₂ enzyme family comprises both calcium-dependent and calcium-independent forms in the CNS. The small-molecular-mass, secreted PLA₂s (sPLA₂s) are induced in systemic inflammatory responses and localized inflammation¹³. The large-molecular-mass, intracellular, cytosolic PLA₂ (cPLA₂), though expressed almost ubiquitously, is regulated both transcriptionally and at the activity level by inflammatory stimuli¹⁴. We assayed PLA₂ activities in total protein extracts from the paw, spinal cord and brain of naive and inflamed rats (12 h) using assay systems favouring either sPLA₂ (Fig. 2a) or cPLA₂ (Fig. 2b). No increase in PLA₂ activity in the spinal cord or brain was produced by paw inflammation. As expected, there were significant increases in both types of activity in the inflamed paw (Fig. 2a, b). Western blot and immunohistochemical analyses of cPLA₂ also failed to show protein induction within the spinal cord (Fig. 2c and data not shown). The basal level and activity of central PLA₂s are sufficient, therefore, to support increased eicosanoid production in the CNS after peripheral inflammation and central Cox-2 induction. In contrast to the peripheral site of inflammation where a two-component (Cox-2 and PLA₂) system for prostanoid release occurs¹⁵, Cox-2 alone appears to be pivotal in central PGE₂ induction. Although both sPLA₂ and cPLA₂ are present in rat spinal cord, our studies do not show which PLA₂ forms couple to Cox-2 for central prostanoid production.

Two general mechanisms may contribute to central Cox-2 induction in response to peripheral inflammation: sensory inflow from nerve fibres innervating the inflamed hindpaw may produce a

synaptic activity-dependent increase in transcription, and circulating pro-inflammatory cytokines may enhance Cox-2 transcription. The delay in transcription argues in favour of a slow, humoral signal; but to investigate whether primary sensory neural activity is involved in central Cox-2 induction, we blocked sciatic nerve conduction before initiating peripheral inflammation using an injectable suspension of bupivacaine-polyester microspheres. Complete sensory and motor blockade of the sciatic nerve for more than 24 h reduced (to 5 ± 1.5 -fold, $n = 3$, Fig. 3a), but did not eliminate, Cox-2 mRNA induction in the lumbar spinal cord or PGE₂ levels in the CSF after hindpaw inflammation (Fig. 3b). Electrical stimulation of the sciatic nerve at C-fibre strength for 30 min did increase Cox-2 in the spinal cord (6 h later), but by a much smaller extent than inflammation (less than threefold, data not shown). These data, together with the widespread distribution of Cox-2 induction in the CNS, point to a non-neuronal, activity-independent factor as a major inducer of central Cox-2 expression after peripheral inflammation. The pro-inflammatory cytokine

IL-1 β is a prime candidate, considering both its induction of Cox-2 in non-neuronal cells⁶ and its upregulation following inflammation¹⁶.

A massive upregulation of IL-1 β (more than 10,000-fold) occurs in the inflamed paw soon after CFA administration and lasts for several days (Fig. 3c). IL-1 β is also increased 50 and 20-fold in the CSF 2 and 4 h after inflammation, respectively (Fig. 3c), which precedes the peak upregulation of Cox-2 mRNA. The type-I IL-1 β receptor is strongly expressed in the spinal cord, especially in laminae I–III of the dorsal horn (Fig. 3d), where Cox-2 is induced upon inflammation (Fig. 1c). The failure of dorsal rhizotomy to affect IL-1 receptor expression (data not shown) suggests that this receptor is expressed by intrinsic spinal cord cells. To study whether IL-1 β induces central Cox-2, rats were subjected to intravenous and intrathecal administration of IL-1 β , and Cox-2 mRNA in the lumbar spinal cord was measured 5 h later. A large intravenous dose of IL-1 β (1 μ g) upregulated spinal Cox-2 mRNA only fourfold (Fig. 3e), which is much less than peripheral inflammation

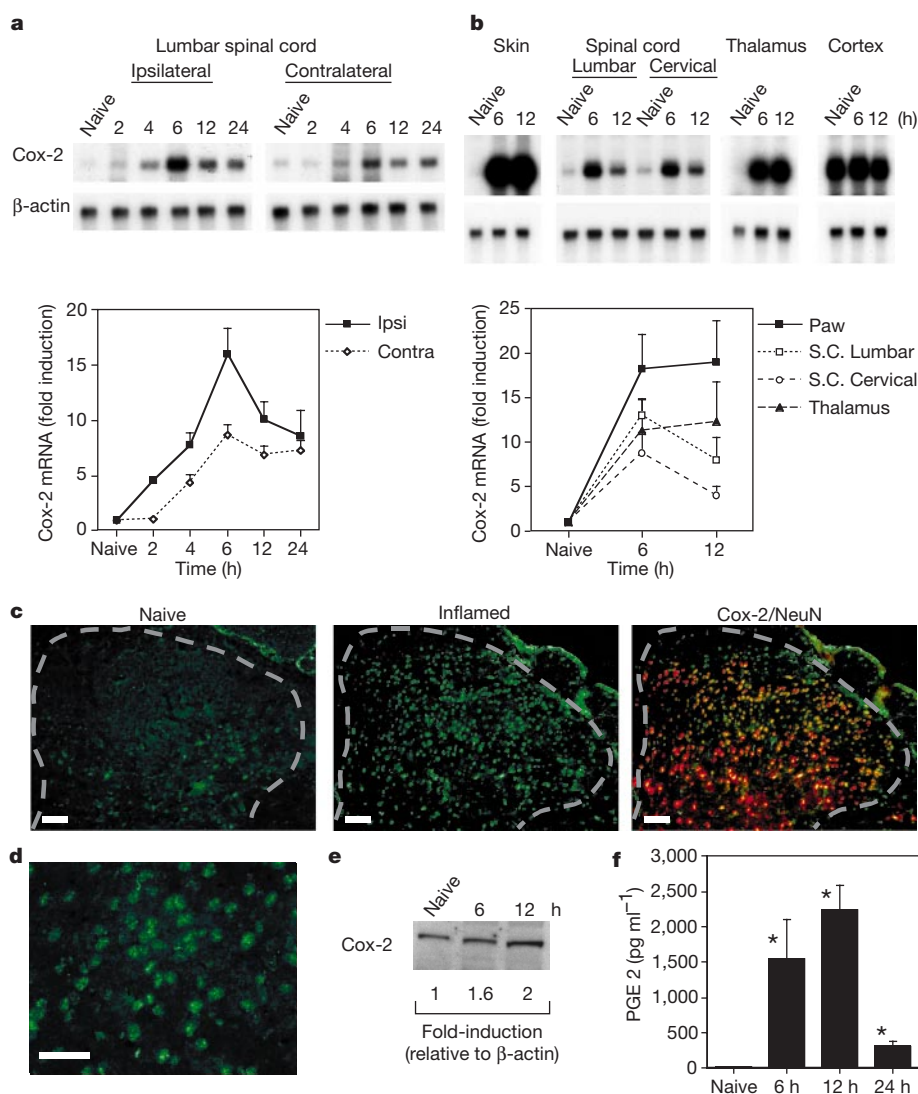


Figure 1 Cox-2 induction and PGE₂ release in the CNS. **a, b**, Cox-2 mRNA induction after unilateral hindpaw inflammation. Increased Cox-2 mRNA in ipsi- and contralateral rat lumbar L4/L5 spinal cord, inflamed skin, lumbar and cervical spinal cord, thalamus and cortex at different times after intraplantar CFA, detected by RNase protection assay with β -actin mRNA as loading control. The graphs show the relative values of Cox-2 mRNA induction with respect to naive rats, after normalizing for β -actin mRNA levels ($n = 3$ experiments, mean $\pm$ s.e.m.). **c**, Cox-2 expression in ipsilateral lumbar dorsal horn

neurons detected immunohistochemically in naive rats and following inflammation (12 h). Most Cox-2-labelled profiles are neuronal (double label with neuronal marker NeuN). **d**, Perinuclear localization of Cox-2 in dorsal horn neurons. Scale bars, 50 μ m. **e**, Western blot analysis of Cox-2 expression in the spinal cord of naive, 6 h and 12 h inflamed rats, with β -actin as loading control. Two independent experiments were analysed. **f**, Increase in PGE₂ in CSF after CFA-induced inflammation. Asterisk, $P < 0.002$ when compared to naive; $n = 6$ –7 rats per time point.

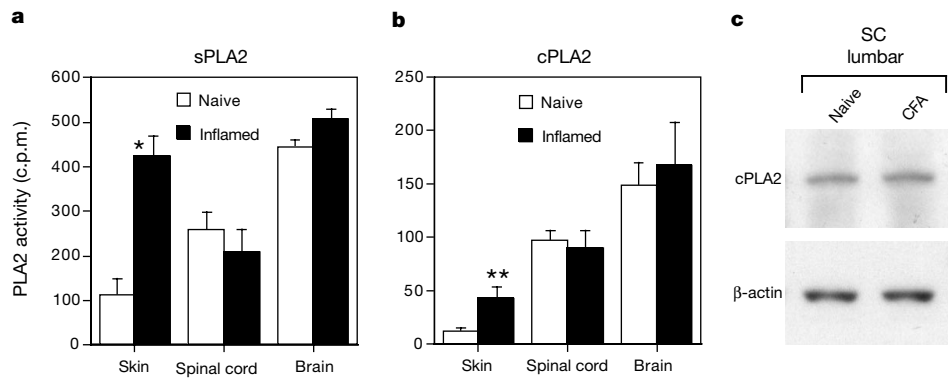


Figure 2 PLA₂ activity in the CNS is not regulated by inflammation. **a**, **b**, sPLA₂ and cPLA₂-like activity assayed from fresh rat skin, spinal cord and whole-brain homogenates from naive and inflamed (12 h) animals. The PLA₂ activity of the indicated tissue homogenate (30 μg) is expressed as c.p.m. of [¹⁴C]arachidonic acid released from [¹⁴C]2-arachidonyl-phosphatidylcholine in 30 min (**a**) or [¹⁴C]2-arachidonyl-phosphatidylethanolamine in

120 min (**b**). Asterisk, $P < 0.01$ and double asterisk, $P < 0.05$ when compared to naive, Student's *t*-test, $n = 4$ rats. **c**, Western blot analysis of cPLA₂ expression in naive and 12 h inflamed rats. β-actin was used as loading control (the same observation was made in tissue from three other rats).

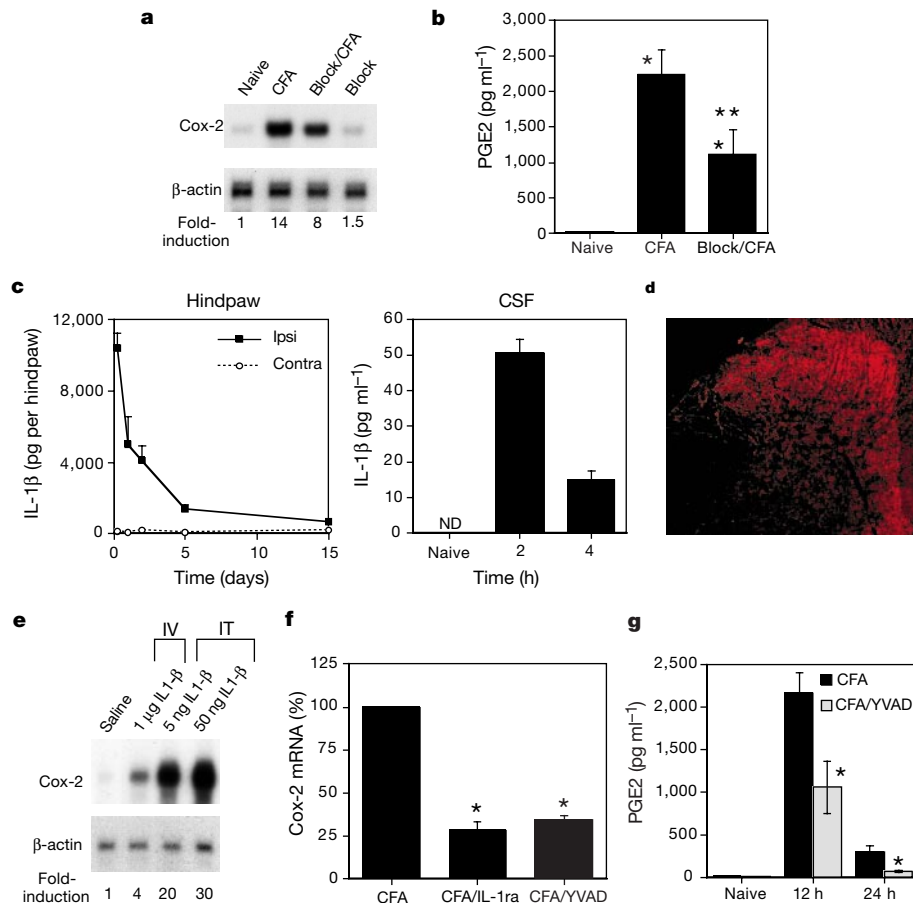


Figure 3 Afferent nerve activity and IL-1β-dependent Cox-2 induction. **a**, Sciatic nerve blockade by perineural bupivacaine treatment sufficient to abolish all afferent activity reduces but does not eliminate the CFA-induced Cox-2 mRNA increase 6 h after inflammation. Fold induction is expressed with respect to naive, after normalizing to β-actin mRNA levels. **b**, Increase in PGE₂ levels in the CSF of inflamed rats (6 h; $n = 13$) and in inflamed rats with sciatic nerve blockade (Block/CFA; $n = 6$). Asterisk, $P < 0.01$ when compared to naive, double asterisk, $P < 0.05$ when compared to CFA. **c**, Increase in IL-1β levels in inflamed hindpaw skin (ipsi) compared with contralateral hindpaw (contra; $n = 4$). IL-1β levels in the CSF 2 and 4 h following intraplantar CFA in the hindpaw

(ND, not detected < 2 pg ml⁻¹, $n = 4$). **d**, Immunolocalization of IL-1 receptor type I in dorsal horn. Scale, 100 μm. **e**, Intrathecal (IT), and to a lesser extent intravenous (IV), IL-1β administration increases Cox-2 mRNA in spinal cord relative to saline-injected animals. **f**, Effect of IT administration, 15 min before intraplantar CFA injection, of 6 μg of IL-1 receptor antagonist (CFA/IL-1ra) and 1 nmole of ICE inhibitor YVAD (CFA/YVAD) on Cox-2 mRNA expression 6 h later. Asterisk, $P < 0.005$ when compared to CFA, $n = 3$. **g**, YVAD decreased PGE₂ in the CSF by 50 and 70% 12 and 24 h after inflammation, respectively. Asterisk, $P < 0.01$ when compared to CFA, $n = 7-10$ per group.

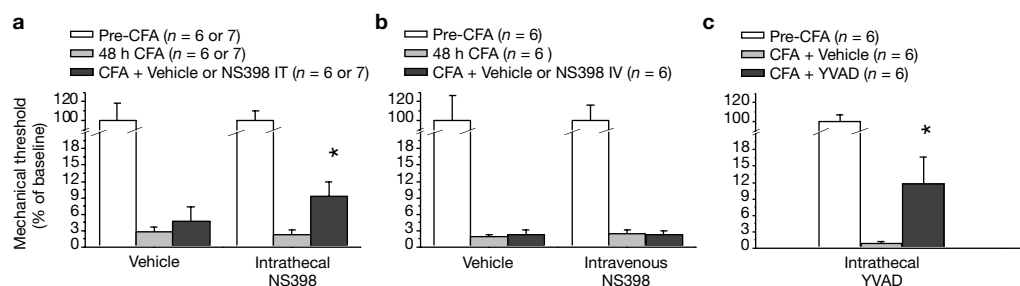


Figure 4 Intrathecal but not intravenous administration of 30 μ g of the Cox-2 inhibitor NS398 48 h after induction of CFA inflammation significantly attenuates mechanical hypersensitivity. **a**, **b**, Intrathecal (**a**) but not intravenous (**b**) administration of NS398 shifted the mechanical threshold from 0.2 ± 0.07 to 0.83 ± 0.23 g. **c**, The ICE inhibitor

YVAD has a similar effect. Intrathecal administration of YVAD shifted the mechanical threshold from 0.15 ± 0.05 to 1.9 ± 0.8 g. Values were normalized to the pre-CFA baseline of each group. Asterisk, $P < 0.05$.

(16-fold). A much greater effect (20- and 30-fold) was produced by intrathecal injection of lower IL-1 β doses (5 and 50 ng, respectively) (Fig. 3e). Even though levels in the inflamed tissue were very high (Fig. 3c), plasma IL-1 β levels did not increase significantly over basal levels (158 ± 25 pg ml $^{-1}$) at 2, 4 or 6 h after inflammation (122 ± 16 , 181 ± 47 , 313 ± 158 ; $n = 8$). As circulating IL-1 β is unlikely to elevate CSF IL-1 β levels, another signal, such as IL-6, must be responsible¹⁷. Whether central IL-1 β is produced by endothelial, glial or neuronal cells needs to be determined. Central PGE $_2$ levels were significantly upregulated (50-fold) following intrathecal administration of IL-1 β (5 ng) to $1,204 \pm 360$ pg ml $^{-1}$ from naive levels of 24 ± 10 pg ml $^{-1}$ ($P < 0.005$, $n = 6$), showing that central IL-1 β -mediated Cox-2 induction leads to central prostanoid production.

Intrathecal administration of recombinant IL-1 β receptor antagonist (IL-1 ra; 6 μ g) 30 min before CFA injection, to neutralize central IL-1 β signalling, reduced Cox-2 mRNA levels by 75% 6 h after inflammation (Fig. 3f). Caspase-1 (also known as interleukin-1 β -converting enzyme or ICE) converts pro-interleukin-1 β to the active form of IL-1 β . A 30-min intrathecal pretreatment with the ICE inhibitor YVAD (1 nmole, 0.5 μ g) blocked the induction of spinal Cox-2 mRNA by 65% 6 h after peripheral inflammation (Fig. 3f). Intrathecal pretreatment with YVAD (1 nmole) also reduced central PGE $_2$ by 50% at 12 and 24 h after inflammation, respectively (Fig. 3g). These data show that spinal IL-1 β is responsible, perhaps in conjunction with other cytokines such as IL-6 (ref. 17) or IL-18, which is also converted to an active form by ICE¹⁸, for central transcriptional activation of Cox-2 after peripheral inflammation, and that Cox-2 induction is the major limiting factor in the production or release (or both) of central PGE $_2$.

We used two approaches to test the behavioural consequences of central Cox-2 induction: intrathecal administration of either the ICE inhibitor YVAD to inhibit local IL-1 β production or a selective Cox-2 inhibitor¹⁹. Intrathecal, but not intravenous, administration of a low dose of the Cox-2 antagonist NS398 (30 μ g) significantly reduced (4.2-fold) mechanical hyperalgesia 48 h after peripheral CFA-induced inflammation (Fig. 4a, b). YVAD (1 nmole) had a similar effect (12-fold) on mechanical hypersensitivity (Fig. 4c). Neither inhibitor had a significant effect on mechanical or thermal pain sensitivity in naive rats (data not shown). Blocking central prostanoid production significantly attenuates, but does not abolish, the mechanical hypersensitivity produced by peripheral inflammation, unlike high-dose intrathecal morphine²⁰.

Prostanoid production in the spinal cord²¹ following Cox-2 induction in dorsal horn neurons is likely to contribute to the establishment and maintenance of peripheral inflammatory hypersensitivity by facilitating transmitter release from nociceptive fibres²² and by direct activation of prostanoid receptors on dorsal horn neurons²³. The heightened pain sensitivity immediately adjacent to inflamed tissue is likely to represent an additive effect of

synaptic neuromodulators released from the central terminals of sensory fibres innervating the inflamed tissue, such as glutamate and substance P, and prostanoids produced by dorsal horn neurons, which together can produce central sensitization²⁴. The rapid, substantial and widespread changes in Cox-2 and ensuing prostanoid production in the CNS, even after localized inflammation, may alter many neural functions apart from pain sensitivity, and may contribute to the generalized illness behavioural syndrome of infectious diseases, which manifests as fever, lethargy and anorexia. IL-1 ra administration or a null mutation of ICE reduces such sickness behaviour^{25,26}. Optimal analgesic therapy with Cox-2 inhibitors should, therefore, include targeting central Cox-2 by promoting brain penetration. Preventing transcriptional upregulation of Cox-2 gene expression in the CNS by inhibiting central IL-1 β upregulation, blocking IL-1 β receptors or their signal transduction pathways or the use of centrally acting Cox-2 inhibitors and prostanoid receptor antagonists, might represent diverse means of normalizing pain sensitivity after local inflammation or systemic infection. □

Methods

RNAse protection

Intraplantar CFA was injected into the rat left hindpaw under halothane (2%) anaesthesia. Rats were killed and dissected at the indicated time points and RNAse protection assays were carried out on total RNA extracts using the RPA III Kit (Ambion). The template for a Cox-2 radiolabelled riboprobe was generated by PCR from rat dorsal root ganglion complementary DNA using primers 5'-GCAATCCTGTGCTGTCCAAACCA-3' and 5'-TTGGGATCCGGGATGAACCTCT-3', and subsequently cloned into pCRII vector (Invitrogen). Fold change was determined with respect to β -actin loading controls and at least three independent measurements were made for each observation, one rat per time point per measurement.

Immunohistochemistry, western blot and ELISA

Transverse spinal cord sections (20 μ m) were cut and processed for immunohistochemistry using Cox-2 (Santa Cruz), NeuN (Chemicon) and IL-1 receptor type I (Research Diagnostics) antibodies as described²⁷. Four independent animals were studied for each. 100–200 μ l of CSF was withdrawn from the cisterna magna of anaesthetized rats using a 27G needle; PGE $_2$ and PGD $_2$ levels were studied using a prostaglandin E2 enzyme-immunoassay (Amersham Pharmacia Biotech) and prostaglandin D2 enzyme-immunoassay (Cayman Chemical). IL-1 β levels were measured with an ELISA¹⁶. Western blot analysis of Cox-2 expression in total spinal cord protein extracts was performed using a goat Cox-2 polyclonal antibody (Santa Cruz).

Sciatic nerve block

A suspension of biodegradable bupivacaine-polyester microspheres²⁸ was injected 30 min before the CFA injection and was sufficient to produce complete sensory and motor blockade of the sciatic nerve for 48 h as assessed behaviourally, and as has been shown electrophysiologically²⁸.

PLA2 activity assay and immunodetection

Rat organs were Dounce homogenized in ice cold buffer²⁹ containing protease inhibitors and 30 μ g of fresh total organ homogenate were assayed for cPLA $_2$ and sPLA $_2$ activity²⁹. Immunodetection of cPLA $_2$ was performed using a polyclonal antibody raised against the first 129 amino-acid residues (a gift from A. Cybulsky) as reported³⁰.

Mechanical hypersensitivity

Mechanical hypersensitivity was assessed in naive and inflamed rats using calibrated von Frey filaments (0.017–95.5) as described¹⁶.

Statistics

Either Student's *t* or Mann–Whitney rank sum tests were used.

Received 17 October 2000; accepted 5 January 2001.

1. Vane, J. R., Bakhle, Y. S. & Botting, R. M. Cyclooxygenases 1 and 2. *Annu. Rev. Pharmacol. Toxicol.* **38**, 97–120 (1998).
2. McCleskey, E. W. & Gold, M. S. Ion channels of nociception. *Annu. Rev. Physiol.* **61**, 835–856 (1999).
3. Woolf, C. J. & Salter, M. W. Neuronal plasticity-increasing the gain in pain. *Science* **288**, 1765–1768 (2000).
4. Dantzer, R. *et al.* Molecular basis of sickness behavior. *Ann. NY Acad. Sci.* **856**, 132–138 (1998).
5. O'Banion, M. K. Cyclooxygenase-2: molecular biology, pharmacology, and neurobiology. *Crit. Rev. Neurobiol.* **13**, 45–82 (1999).
6. Maier, J. A. M., Hlas, T. & Maciag, T. Cyclooxygenase is an immediate early gene induced by interleukin-1 in human endothelial cells. *J. Biol. Chem.* **265**, 10805–10808 (1990).
7. Muth-Selbach, U. S., Tegeder, L., Brune, K. & Geisslinger, G. Acetaminophen inhibits spinal prostaglandin E₂ release after peripheral noxious stimulation. *Anesthesiology* **91**, 231–239 (1999).
8. Smith, C. J. *et al.* Pharmacological analysis of cyclooxygenase-1 in inflammation. *Proc. Natl Acad. Sci. USA* **95**, 13313–13318 (1998).
9. Minami, T. *et al.* Characterization of EP-receptor subtypes involved in allodynia and hyperalgesia induced by intrathecal administration of prostaglandin E₂ to mice. *Br. J. Pharmacol.* **112**, 735–740 (1994).
10. Beiche, F., Scheuerer, S., Brune, K., Geisslinger, G. & Goppelt-Strube, M. Up-regulation of cyclooxygenase-2 mRNA in the rat spinal cord following peripheral inflammation. *FEBS Lett.* **390**, 165–169 (1996).
11. Hay, C. H., Trevethick, M. A., Wheelton, A., Bowers, J. S. & de Belleruche, J. S. The potential role of spinal cord cyclooxygenase-2 in the development of Freund's complete adjuvant-induced changes in hyperalgesia and allodynia. *Neuroscience* **78**, 843–850 (1997).
12. Beiche, F., Brune, K., Geisslinger, G. & Goppelt-Strube, M. Expression of cyclooxygenase isoforms in the rat spinal cord and their regulation during adjuvant-induced arthritis. *Inflamm. Res.* **47**, 482–487 (1998).
13. Hietaranta, A. *et al.* Extracellular phospholipase A₂ in relation to systemic inflammatory response syndrome (SIRS) and systemic complications in severe acute pancreatitis. *Pancreas* **18**, 385–391 (1999).
14. Levy, R., Lowenthal, A. & Dana, R. Cytosolic phospholipase A₂ is required for the activation of the NADPH oxidase associated H⁺ channel in phagocyte-like. *Adv. Exp. Med. Biol.* **479**, 125–135 (2000).
15. Saunders, M. A. *et al.* Mechanisms of prostaglandin E₂ release by intact cells expressing cyclooxygenase-2: evidence for a 'two-component' model. *J. Pharmacol. Exp. Ther.* **288**, 1101–1106 (1999).
16. Safieh-Garabedian, B., Poole, S., Allchorne, A., Winter, J. & Woolf, C. J. Contribution of interleukin-1 β to the inflammation-induced increase in nerve growth factor levels and inflammatory hyperalgesia. *Br. J. Pharmacol.* **115**, 1265–1275 (1995).
17. Cartmell, T., Poole, S., Turnbull, A. V., Rothwell, N. J. & Luheshi, G. N. Circulating interleukin-6 mediates the febrile response to localised inflammation in rats. *J. Physiol.* **526**, 653–661 (2000).
18. Fantuzzi, G. & Dinarello, C. A. Interleukin-18 and interleukin-1 beta: two cytokine substrates for ICE (caspase-1). *J. Clin. Immunol.* **19**, 1–11 (1999).
19. Lashbrook, J. M. *et al.* Synergistic antiallodynic effects of spinal morphine with ketorolac and selective COX-1 and COX-2-inhibitors in nerve-injured rats. *Pain* **82**, 65–72 (1999).
20. Yaksh, T. L. & Rudy, T. A. Analgesia mediated by a direct spinal action of narcotics. *Science* **192**, 1357–1358 (1976).
21. Dirig, D. M., Konin, G. P., Isakson, P. C. & Yaksh, T. L. Effect of spinal cyclooxygenase inhibitors in rat using the formalin test and in vitro prostaglandin E₂ release. *Eur. J. Pharmacol.* **331**, 155–160 (1997).
22. Southall, M. D., Michael, R. L. & Vasko, M. R. Intrathecal NSAIDs attenuate inflammation-induced neuropeptide release from spinal cord slices. *Pain* **78**, 39–48 (1998).
23. Yamamoto, T. & Sakashita, Y. COX-2 inhibitor prevents the development of hyperalgesia induced by intrathecal NMDA or AMPA. *NeuroReport* **9**, 3869–3873 (1998).
24. Malmberg, A. B. & Yaksh, T. L. Hyperalgesia mediated by spinal glutamate or substance P receptor blocked by spinal cyclooxygenase inhibition. *Science* **257**, 1276–1279 (1992).
25. Swiergiel, A. H., Smagin, G. N., Johnson, L. J. & Dunn, A. J. The role of cytokines in the behavioral responses to endotoxin and influenza virus infection in mice: effects of acute and chronic administration of the interleukin-1-receptor antagonist (IL-1ra). *Brain Res.* **776**, 96–104 (1997).
26. Yao, J. H. *et al.* Mice deficient in interleukin-1 β converting enzyme resist anorexia induced by central lipopolysaccharide. *Am. J. Physiol.* **277**, R1435–R1443 (1999).
27. Amaya, E. *et al.* Diversity of expression of the sensory neuron specific TTX-resistant voltage-gated sodium ion channels SNS and SNS2. *Mol. Cell. Neurosci.* **15**, 331–342 (2000).
28. Curley, J. *et al.* Injectable biodegradable bupivacaine/polyester microspheres. *Anesthesiology* **84**, 1401–1410 (1996).
29. Bingham, C. O. *et al.* Low molecular weight group IIA and group V phospholipase A₂ (PLA₂) enzymes have different intracellular locations in mouse bone marrow-derived mast cells. *J. Biol. Chem.* **274**, 31476–31484 (1999).
30. Sapirstein, A., Spech, R. A., Witzgall, R. & Bonventre, J. V. Cytosolic phospholipase A₂ (PLA₂), but not secretory PLA₂, potentiates hydrogen peroxide cytotoxicity in kidney epithelial cells. *J. Biol. Chem.* **271**, 21505–21513 (1996).

Acknowledgements

We thank H. Baba, C. Van der Meer, M. Costigan, R.-R. Ji, K. Befort, D. Borsook and L. Kobierski for assistance and advice and C. Berde for the microspheres. Financial support was provided by NIH and Pfizer (C.J.W.). Cytokine reagents generated by EC CYBRAINET Biomed II and TMR Programme (S.P.).

Correspondence and requests for materials should be addressed to C.J.W. (e-mail: woolf.clifford@mgh.harvard.edu).

Homologues of Twisted gastrulation are extracellular cofactors in antagonism of BMP signalling

Ian C. Scott^{*†}, Ira L. Blitz^{‡†}, William N. Pappano^{§†}, Sarah A. Maas[§], Ken W. Y. Cho[‡] & Daniel S. Greenspan^{*§}

^{*} Department of Pathology and Laboratory Medicine, [§] Department of Biomolecular Chemistry, University of Wisconsin Medical School, 1300 University Avenue, Madison, Wisconsin 53706, USA

[‡] Department of Developmental and Cell Biology, and the Developmental Biology Center, University of California, Irvine, California 92697, USA

[†] These authors contributed equally to this work

Twisted gastrulation (TSG) is involved in specifying the dorsal-most cell fate in *Drosophila* embryos¹, but its mechanism of action is poorly understood. TSG has been proposed to modify the action of Short gastrulation (SOG), thereby increasing signalling by the bone morphogenetic protein (BMP) Decapentaplegic. SOG, an inhibitor of BMP signalling, is in turn inactivated by the protease Tolloid^{2,3}. Here we identify Tsg gene products from human, mouse, *Xenopus*, zebrafish and chick. Expression patterns in mouse and *Xenopus* embryos are consistent with *in vivo* interactions between Tsg, BMPs and the vertebrate SOG orthologue, chordin. We show that Tsg binds both the vertebrate Decapentaplegic orthologue BMP4 and chordin, and that these interactions have multiple effects. Tsg increases chordin's binding of BMP4, potentiates chordin's ability to induce secondary axes in *Xenopus* embryos, and enhances chordin cleavage by vertebrate tolloid-related proteases at a site poorly used in Tsg's absence; also, the presence of Tsg enhances the secondary axis-inducing activity of two products of chordin cleavage. We conclude that Tsg acts as a cofactor in chordin's antagonism of BMP signalling.

To identify and characterize possible vertebrate homologues of *Drosophila* TSG, we used RACE (rapid amplification of complementary DNA ends)-PCR, ESTs (expressed sequence tags) and low-stringency hybridizations to obtain full-length sequences from human, mouse, *Xenopus*, zebrafish and chick (see Supplementary Information). We mapped Tsg genes to human chromosome 18p11, the syntenic distal portion of mouse chromosome 17, and 63.9–75.2 centimorgans (cM) from the top of zebrafish linkage group 2.

We found Tsg to be coexpressed with chordin and various BMPs in vertebrate development. In *Xenopus*, maternal Tsg RNA was detected in eggs by RT (reverse transcription)-PCR (Fig. 1a), while whole-mount *in situ* hybridization showed uniform Tsg expression across the entire animal hemisphere and marginal zone of the early gastrula (Fig. 1b). At tailbud stage, Tsg, chordin and BMP4 expression domains partially overlap in the developing tail, anterior brain, eye and heart (data not shown; see also ref. 4). In mouse, Tsg is broadly expressed throughout the 7.5-days-post-coitus (7.5-d.p.c.) gastrula and in extraembryonic tissues (data not shown). Chordin, Tsg and BMPs 2, 4 and 7 are highly expressed in the digital rays of 15.5- (data not shown) and 17.5-d.p.c. embryo hindlimbs (Fig. 1c–h). Strong chordin expression in the interzone of the joint cavity is juxtaposed with strong Tsg expression at the joint articular surfaces and the interzone. Thus, Tsg is properly situated for potential interactions with chordin and BMPs during various stages of vertebrate embryogenesis.

In *Drosophila*, TSG influences cleavage of the chordin orthologue Short gastrulation (SOG) by Tolloid, altering the pattern of SOG cleavage products⁵. There are four mammalian tolloid-related proteases⁶. Two of these, BMP1 and mammalian tolloid-like 1 (mTll1), each cleave chordin at two specific sites, yielding

fragments of relative molecular mass (M_r) $\sim 15K$, $\sim 13K$ and $\sim 83K$, corresponding to the amino-terminal, carboxy-terminal, and internal portions of chordin, respectively (ref. 6 and Fig. 2c, top). Murine Tsg appears to enhance cleavage of mouse chordin (Fig. 2a) and to influence the relative abundance of cleavage products, such that fragments of $M_r \sim 65K$ and $\sim 29K$, minor forms in the absence of Tsg, become major products in the presence of Tsg (Fig. 2b, c bottom). A third related protease, mammalian tolloid (mTld), which has little detectable chordin-cleaving activity⁶, has significant activity in the presence of Tsg, also producing the $\sim 65K$ and $\sim 29K$ fragments as major forms (data not shown). The fourth mammalian tolloid-like protease, mTll2, lacks chordin-processing activity in the presence (data not shown) or absence⁶ of Tsg.

The 65K and 29K chordin cleavage products preferentially produced in the presence of Tsg are subfragments of the 83K internal chordin fragment (Fig. 2c), as established by N-terminal amino-acid sequencing, and result from cleavage at a previously unmapped site between Ala 670 and Thr 671. Thus, the 29K form contains chordin cysteine-rich repeats (CRs) 2 and 3, whereas the 65K form contains no CR domains.

To determine how Tsg affects chordin cleavage, we first examined Tsg's ability to physically interact with tolloid-like proteases. Co-immunoprecipitation of Tsg with BMP1 or mTll1 failed to detect physical interactions (Fig. 2d). However, co-immunoprecipitations showed that Tsg binds chordin (Fig. 2e).

We next examined whether Tsg/chordin interactions might influence chordin's ability to bind BMP4. Co-immunoprecipitation of chordin and BMP4 is greatly enhanced in the presence of Tsg (Fig.

2f). We also found that Tsg binds BMP4 (Fig. 2g, h). In summary, Tsg's interactions with chordin and/or BMP4 enhance chordin/BMP4 complex formation, suggesting that Tsg might enhance chordin's antagonism of BMP signalling.

We examined whether Tsg could influence chordin inhibition of BMP signalling in early *Xenopus* embryos. Microinjection of Tsg messenger RNA alone yields a dose-dependent reduction in length of the anterior–posterior axis and slight narrowing of the posterior in early tailbud stages (Fig. 3a). At high doses, the head is also slightly reduced and the tail is more strongly kinked. These embryos, when developed to a later stage, appear indistinguishable from those obtained under similar conditions in ref. 7). In that study, molecular marker analysis also shows that Tsg partially dorsalizes ventral marginal zone explants (also see below) and has late effects on patterning of prechordal mesoderm and anterior endoderm. Posterior defects induced by Tsg overexpression may arise through alteration of BMP4 or chordin function in the developing tailbud region.

Ventral inhibition of BMP signalling in *Xenopus* gastrulae by chordin overexpression results in development of secondary dorsal axes^{8,9}. Thus, we predicted that Tsg's ability to increase chordin/BMP4 complex formation might result in potentiation of chordin's axis-inducing activity. We first identified a dose of chordin mRNA that inefficiently induces axes when ventrally injected. Three picograms of chordin mRNA results in efficient axis induction, whereas 1 pg failed to induce axes. However, co-injection of 1 pg chordin mRNA with 45 pg Tsg mRNA efficiently restored secondary axis formation (Fig. 3b). Ventral injection with 50 pg Tsg mRNA alone

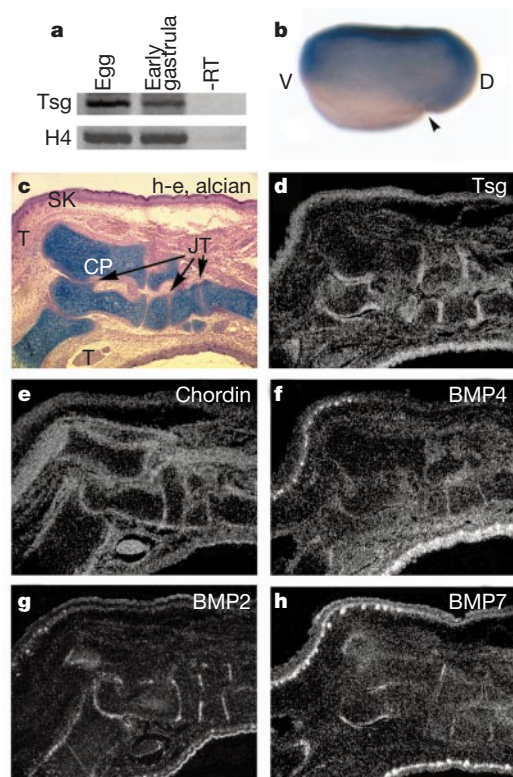


Figure 1 Tsg expression patterns in *Xenopus* and mouse embryos. **a**, RT–PCR shows that *Xenopus* Tsg is expressed both maternally and during gastrulation. **b**, Whole-mount *in situ* hybridization of a stage-10.25 gastrula shows uniform *Xenopus* Tsg expression in the animal half. **c–h**, Serial sections of 17.5-d.p.c. mouse hindlimbs characterized by staining with haematoxylin-eosin and alcian blue (**c**) or by *in situ* hybridization with anti-sense Tsg (**d**), chordin (**e**), BMP2 (**f**), -4 (**g**), or -7 (**h**) riboprobes, showing overlapping expression patterns. SK, skin; T, tendon; CP, cartilage primordia; JT, joint.

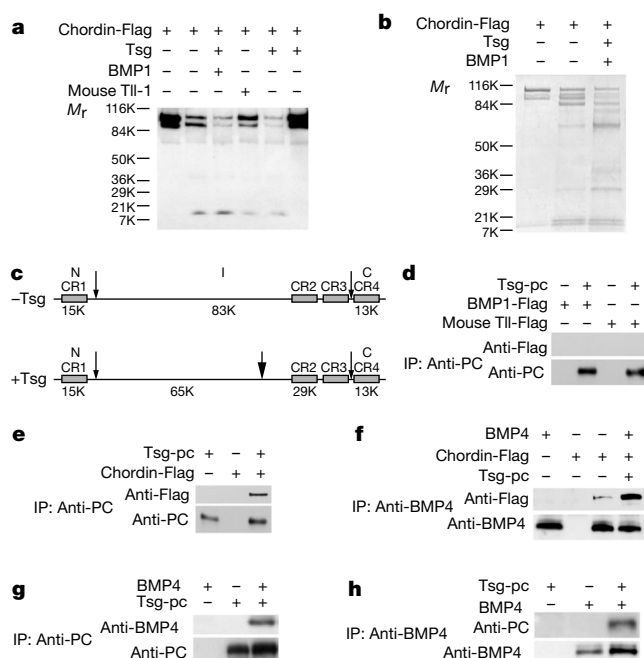


Figure 2 Tsg alters cleavage of chordin by BMP1 and mTll1, and directly interacts with chordin and BMP4 to enhance chordin/BMP4 complex formation. **a**, Immunoblot of Flag-tagged chordin incubated with BMP1 or mTll1 in the presence (+)/absence (–) of Tsg. Bands were detected using anti-Flag antibodies. **b**, Electrophoretic patterns, visualized by zinc staining⁶, for chordin incubated with BMP1 in the presence/absence of Tsg. **c**, Positions of major chordin cleavage sites (arrows) utilized in the presence/absence of Tsg. **d,e**, Immunoblots of proteins precipitated with anti-protein C antibody (anti-PC) after incubation of (**d**) Flag-tagged BMP1, mTll1, or (**e**) chordin, alone or with protein C-tagged Tsg (Tsg-pc). **f**, Immunoblot of proteins precipitated with anti-BMP4 antibodies (anti-BMP4), after incubation of BMP4 with chordin-Flag with/without Tsg. **g,h**, Immunoblots of proteins immunoprecipitated with (**g**) anti-PC or (**h**) anti-BMP4 antibodies, after incubation of BMP4 and Tsg separately or together.

failed to induce secondary axes (not shown; also, see below). Next, we determined whether the effects produced by co-injection of Tsg and chordin are due to inhibition of BMP signalling. Consistent with this possibility, embryos co-injected with Tsg and chordin mRNAs demonstrate greater reduction of *Xenopus Vent2* expression, a direct target of BMP signalling¹⁰, than embryos injected with either alone (Fig. 3c). Taken together with co-immunoprecipitation

data, these results suggest that Tsg is capable of potentiating chordin/BMP complex formation, thereby inhibiting BMP signalling *in vivo*.

Individual chordin CR1 and CR3 domains retain weak dorsalizing activity in *Xenopus* embryos¹¹. As Tsg enhances chordin's BMP inhibitory activity, we examined whether Tsg might affect the activity of chordin cleavage products. Equimolar amounts of mRNAs corresponding to untagged versions of each of the chordin cleavage products shown in Fig. 2c were microinjected with or without 100 pg Tsg mRNA. Although none of the chordin fragments alone induced axes, both the CR1- and CR2/CR3-containing fragments induced axes when co-injected with Tsg (Fig. 3d–f). No secondary axes or dorsalizing effects were obtained from overexpression of any of the other chordin fragments, even in the presence of Tsg (Fig. 3d–f).

To determine whether Tsg has the ability to form a complex with any of the CR-containing chordin cleavage products, co-immunoprecipitation assays were performed. Both the CR1 and CR2/CR3 fragments, but not CR4, co-immunoprecipitated with Tsg (Fig. 3g). Thus, there is a correlation between those CR-containing fragments which bind Tsg, and those which have enhanced axis-inducing activity in the presence of Tsg. The data suggested that Tsg might increase complex formation between these fragments and BMP4. This issue was examined using antibodies to the CR1-containing fragment in experiments which found co-immunoprecipitation of CR1 and BMP4 to be greatly enhanced in the presence of Tsg (Fig. 3h). As Tsg is capable of potentiating the binding of BMP4 to both full-length chordin and the CR1-containing cleavage product, and Tsg interacts with CR1- and CR2/CR3-containing fragments, the present data indicate that a subset of chordin's CR domains may contain the determinants for Tsg's interaction with full-length chordin.

Models for cell-fate specification along the dorsal–ventral axis of *Drosophila* and vertebrate embryos employ gradients of BMP signalling. In *Drosophila*^{12–17}, a dorsal (high) to ventral (low) gradient of Decapentaplegic (DPP)/Screw (SCW) activity may be shaped by an opposing concentration gradient of SOG, and inhibition of SOG function by Tollid³. However, a paradox exists in this model. The dorsal-most tissue, the amnioserosa, is thought to require peak DPP/SCW signalling, yet loss-of-function mutations in a BMP inhibitor, SOG, lack the amnioserosa^{12–15,18}. According to the model, amnioserosa should instead be expanded in *sog* mutants.

Amnioserosa is similarly absent in *Drosophila tsg* mutants¹. However, recent studies show that expression of certain dorsal marker genes is expanded in both *tsg* and *sog* mutants^{3,5,19}. Additionally, *tsg* mutants can be rescued by expression of a hyperinhibitory SOG fragment ("supersog"), which cannot be inactivated by Tollid⁵. This observation suggests that supersog's BMP-inhibitory activity can replace the function of TSG. These previous findings are consistent with the data presented here and elsewhere^{7,19} suggesting that Tsg homologues can function along with Sog/chordin in BMP inhibition. It has been suggested that Tsg acts as a BMP agonist⁴. Although the data of ref. 19 show *Drosophila* TSG to be a BMP (DPP/SCW) antagonist that inhibits BMP signalling dorsolaterally, they also suggest that TSG promotes BMP signalling at the dorsal midline. Thus, we speculate that Tsg might play dual roles as antagonist and agonist depending on relative concentrations of Tsg, chordin, proteases and BMPs and (or) on as-yet unidentified cofactors. Finally, whereas our study and others^{4,5,11} demonstrate that chordin/SOG cleavage products can modulate BMP signalling, there is as yet no evidence showing roles for these fragments in normal development. In fact, phenotypes observed in *Drosophila sog* and *tsg* mutants can be explained without invoking active roles for SOG cleavage products¹⁹. We conclude that further study is required to resolve various issues generated by data from *Drosophila* embryonic patterning mutants. □

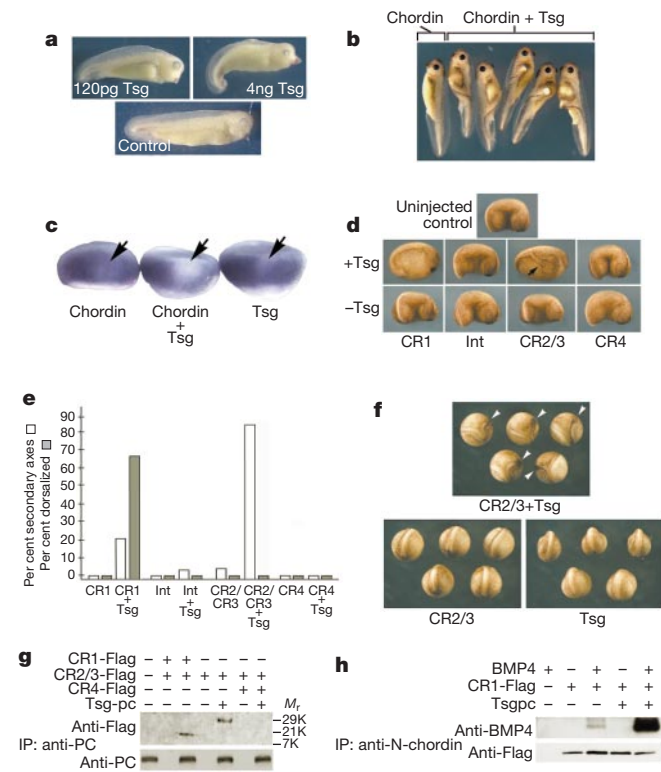


Figure 3 Tsg potentiates secondary axis induction by chordin and chordin fragments and increases their binding of BMP4. **a**, Effects of Tsg overexpression in *Xenopus* embryos. **b**, Whereas 1 pg chordin RNA induced no axes (0/23), co-injection of 1 pg chordin and 45 pg Tsg RNA induced axes in 83% of embryos ($n = 49$). Tsg RNA alone (50 pg) induced no axes ($n = 24$) (not shown). **c**, Whereas 300 fg chordin or 45 pg Tsg RNA alone induced no axes ($n = 25$ each), co-injection of the two induced axes in 80% ($n = 25$) and strong *Xenopus Vent2* inhibition in 75% ($n = 25$) of embryos (arrows show areas of strong or weak *Vent2* inhibition in ventral marginal zones). **d**, Equimolar amounts of CR1 (100 pg), 65K internal 'Int.' (410 pg), CR2/CR3 (150 pg) or CR4 (69 pg) fragment RNA were injected with/without 100 pg Tsg RNA. Without Tsg RNA, no axes were induced (CR1, 0/43; Int., 0/24; CR2/CR3, 1/26; CR4, 0/28) or embryos dorsalized, (CR1, 0/43; Int., 0/24; CR2/CR3, 0/26; CR4, 0/28). In the presence of Tsg, CR2/CR3 efficiently induced axes (84%, $n = 25$), CR1 induced a lesser number (21%, $n = 24$) of small axes, and other fragments inefficiently induced axes (Int., 1/27; CR4, 0/23). CR1+Tsg dorsalized embryos (67%, $n = 24$), including those with secondary axes. **e**, Results of **d** (in a separate experiment, CR1+Tsg induced a higher proportion of secondary axes and lower proportion of dorsalized embryos, relative to CR2/CR3, than in the experiment shown). **f**, Co-injection of 100 pg each Tsg and CR2/CR3 RNA efficiently induced axes (93%, $n = 30$), whereas 100 pg Tsg RNA alone did not induce axes (0%, $n = 25$). 100 pg CR2/CR3 alone inefficiently induced axes (4%, $n = 25$). In all experiments shown in panels **a–f**, untagged versions of CR and Tsg constructs were used. In control experiments, similar to **d–f**, enhancement of axis formation by CR1 in the presence of Tsg is still efficient when RNAs for the two are co-injected with BMP1 RNA (not shown). This excludes the possibility that Tsg enhancement of axis induction in such experiments is through interaction with endogenous full-length chordin, induced via a positive feedback loop¹⁰ by overexpressed chordin fragments. **g**, Immunoblot of proteins precipitated with anti-PC antibodies after incubation of each CR-containing fragment alone or with Tsg-pc. **h**, Immunoblot of proteins precipitated by antibodies (anti-N-chordin) to sequences in the CR1 fragment after incubation of CR1 fragment (1.7 nM) with BMP4 (1.4 nM) with/without Tsg (1.3 nM).

Methods

Isolation of human, mouse, *Xenopus*, zebrafish and chick Tsg sequences

See Supplementary Information.

Gene mapping

Human *TWIST1* was mapped 6.61 centirays (6.61 cR) from WI-5607 and 5.13 cR from D18S464, by radiation hybrid mapping using the GeneBridge 4 panel. Mouse *Twsg1* was mapped with the TJL BSS backcross panel to distal chromosome 17, cosegregating with *Rip1* offset 38 from the centromere. The zebrafish gene was mapped with a zebrafish/hamster radiation hybrid panel (Research Genetics).

Production of recombinant proteins

Mouse Tsg cDNA coding sequences (with/without C-terminal protein C tag), minus signal peptide, were inserted into pCEP-Pu/BM40s²⁰ in frame with the BM40 signal peptide, to maximize secretion. Flag-tagged chordin and proteases were produced as in ref. 6. For Flag-tagged chordin fragments, or corresponding non-tagged mRNAs for *Xenopus* embryo injections, coding sequences were PCR-amplified from full-length mouse chordin cDNA template⁶ (see Supplementary Information for expression construct details). Recombinant proteins were expressed, and Flag-tagged proteins purified, as described⁶. To purify protein C epitope-tagged Tsg (Tsg-pc) (performed at 4 °C), conditioned medium was adjusted to 1 mM CaCl₂, 0.2% Brij-35 and recirculated over an anti-protein C column that was washed and equilibrated per manufacturer's instructions (Boehringer Mannheim). Tsg for enzymatic assays was eluted with 20 mM Tris-HCl, pH 7.5, 0.15 M NaCl, 1 mM CaCl₂, 1 mg ml⁻¹ protein C peptide. Tsg for immunoprecipitations was eluted with 20 mM Tris-HCl, pH 7.5, 0.15 M NaCl, 5 mM EDTA. Protein concentrations were calculated as described⁶.

Immunoblotting and precipitations

Immunoblots employed anti-Flag-BioM2 (Sigma) (1:7,500) plus peroxidase-linked streptavidin (Sigma) (1:5,000); anti-protein C-peroxidase (Boehringer Mannheim) (1:5,000); or anti-BMP4 antisera (R&D) (1:1,000) plus anti-mouse IgG-peroxidase (Amersham) (1:5,000). Amino-terminal sequencing was done as in ref. 6.

Immunoprecipitations used roughly equimolar amounts of purified proteins. In Fig. 2d and e, 150 ng Flag-tagged chordin, BMP1, or mTll1 was incubated with 50 ng Tsg-pc for co-immunoprecipitations (co-IPs) using described conditions²¹. Other co-IPs employed conditions similar to those of ref. 11. For Fig. 2f, 150 ng chordin-Flag was incubated with 50 ng BMP4 (R&D), with/without 50 ng Tsg-pc, followed by addition of monoclonal antibody to BMP4 (R&D) prebound to protein G Sepharose. For Fig. 2g, h, co-IPs of 100 ng each BMP4 and Tsg-pc employed anti-protein C beads (Fig. 2g) or anti-BMP4 antibody prebound to protein G Sepharose (Fig. 2h). For Fig. 3g, co-IPs of Flag-tagged CR2/3 (100 ng), CR1 (50 ng), or CR4 (50 ng) fragment with 100 ng Tsg-pc employed anti-protein C beads. We note that in a related control experiment Tsg was found to co-immunoprecipitate untagged CR1-containing fragment, thus showing that the Flag-tag does not influence binding of Tsg to this fragment (data not shown). For Fig. 3h, co-IP of 50 ng BMP4 with 25 ng Flag-tagged CR1 in the presence/absence of 50 ng Tsg-pc employed rabbit antibodies to peptide EPPALPIRSEKEPLPVRGA, (murine chordin²² residues 30–48), prepared and purified as in ref. 23, prebound to protein A Sepharose. Protein C or Sepharose beads were eluted with Laemmli buffer except in Fig. 3g, in which elution was with 20 mM Tris-HCl, pH 7.5, 150 mM NaCl, 10 mM EDTA.

Chordin cleavage assays

Chordin-Flag (50 ng) was incubated with/without Tsg-pc (50 ng) in buffer A⁶ for 1 h at 37 °C. Twenty nanograms of Flag-tagged BMP1, mTld, mTll1 or mTll2 was then added, followed by incubating for 8 h at 37 °C.

Xenopus embryo manipulations and microinjections

Embryo manipulations were as in ref. 24. Mouse Tsg and chordin mRNAs were prepared from linearized pCS2+ constructs using the SP6 mMessage mMachine kit (Ambion). RNA injections were (Fig. 3a) into the marginal zone of each blastomere at the four-cell stage (30 pg or 1 ng Tsg RNA/blastomere); (Fig. 3b) into the marginal zone of a ventral blastomere (two-cell stage); (Fig. 3c) into one ventral vegetal blastomere at the eight-cell stage; or (Fig. 3d and f) into a single ventral marginal blastomere at the 32-cell stage. Note that injection of CR1 RNA alone into ventral marginal zones at the eight-cell stage dorsalized embryos, as in ref. 11, but that neither CR1 nor CR2/CR3 RNA efficiently dorsalizes when injected at the 32-cell stage. Whole-mount *in situ* hybridization was as in ref. 25 with modifications described in ref. 10, except that pigmented embryos were used in Fig. 3c. For this experiment, embryos were dehydrated in methanol after fixation, rehydrated through a methanol series into PBS-Tween 20, bleached 4 h in PBS-Tween 20 containing 6% H₂O₂, and extensively washed in PBS-Tween 20 before the proteinase K step. Digoxigenin-labelled riboprobes were synthesized using standard methods²⁵.

Murine *in situ* hybridization and histological analysis

Embryos were prepared, riboprobes labelled with [³⁵S]UTP and *in situ* hybridization performed as in ref. 26. Tsg riboprobes were prepared from a linearized Tsg cDNA-pBluescript construct (see Supplementary Information) transcribed with polymerase T7

(antisense) or T3 (sense). Chordin riboprobes were as in ref. 6. BMP2, -4 and -7 riboprobes were as described²⁷. Histological staining was as in ref. 6.

Received 7 August 2000; accepted 5 January 2001.

- Mason, E. D., Konrad, K. D., Webb, C. D. & Marsh, J. L. Dorsal midline fate in *Drosophila* embryos requires *twisted gastrulation*, a gene encoding a secreted protein related to human connective tissue growth factor. *Genes Dev.* **8**, 1489–1501 (1994).
- Francois, V., Solloway, M., O'Neill, J. W., Emery, J. & Bier, E. Dorsal-ventral patterning of the *Drosophila* embryo depends on a putative negative growth factor encoded by the short gastrulation gene. *Genes Dev.* **8**, 2602–2616 (1994).
- Marqués, G. *et al.* Production of a DPP activity gradient in the early *Drosophila* embryo through the opposing actions of the SOG and TLD proteins. *Cell* **91**, 417–426 (1997).
- Oelgeschläger, M., Larrain, J., Geissert, D. & De Robertis, E. M. The evolutionarily conserved BMP-binding protein *Twisted gastrulation* promotes BMP signalling. *Nature* **405**, 757–763 (2000).
- Yu, K. *et al.* Processing of the *Drosophila* Sog protein creates a novel BMP inhibitory activity. *Development* **127**, 2143–2154 (2000).
- Scott, I. C. *et al.* Mammalian BMP-1/ Tolloid-related metalloproteinases, including novel family member mammalian tolloid-like 2, have differential enzymatic activities and distributions of expression relevant to patterning and skeletogenesis. *Dev. Biol.* **213**, 283–300 (1999).
- Chang, C. *et al.* *Twisted gastrulation* can function as a BMP antagonist. *Nature* **410**, 483–487 (2001).
- Sasai, Y. *et al.* *Xenopus chordin*: A novel dorsalizing factor activated by organizer-specific homeobox genes. *Cell* **79**, 779–790 (1994).
- Piccolo, S., Sasai, Y., Lu, B. & De Robertis, E. M. Dorsal-ventral patterning in *Xenopus*: Inhibition of ventral signals by direct binding of Chordin to BMP-4. *Cell* **86**, 589–598 (1996).
- Blitz, I. L., Shimmi, O., Wünnenberg-Stapleton, K., O'Connor, M. B. & Cho, K. W. Y. Is Chordin a long-range or short-range acting factor? Roles for BMP1-related metalloproteinases in regulating chordin action. *Dev. Biol.* **223**, 120–138 (2000).
- Larrain, J. *et al.* BMP-binding modules in chordin: a model for signalling regulation in the extracellular space. *Development* **127**, 821–830 (2000).
- Arora, K., Levine, M. S. & O'Connor, M. B. The screw gene encodes a ubiquitously expressed member of the TGF- β family required for specification of dorsal cell fates in the *Drosophila* embryo. *Genes Dev.* **8**, 2588–2601 (1994).
- Ferguson, E. L. & Anderson, K. V. Localized enhancement and repression of the activity of the TGF- β family member, *decapentaplegic*, is necessary for dorsal-ventral pattern formation in the *Drosophila* embryo. *Development* **114**, 583–597 (1992).
- Arora, K. & Nüsslein-Volhard, C. Altered mitotic domains reveal fate map changes in *Drosophila* embryos mutant for zygotic dorsoventral patterning genes. *Development* **114**, 1003–1024 (1992).
- Wharton, K. A., Ray, R. P. & Gelbart, W. M. An activity gradient of *decapentaplegic* is necessary for the specification of dorsal pattern elements in the *Drosophila* embryo. *Development* **117**, 807–822 (1993).
- Holley, S. A. *et al.* The *Xenopus* dorsalizing factor noggin ventralizes *Drosophila* embryos by preventing DPP from activating its receptor. *Cell* **86**, 607–617 (1996).
- Holley, S. A. *et al.* A conserved system for dorsal–ventral patterning in insects and vertebrates involving *sog* and *chordin*. *Nature* **376**, 249–253 (1995).
- Zusman, S. B., Sweeton, D. & Wieschaus, E. F. *Short gastrulation*, a mutation causing delays in stage specific cell shape changes during gastrulation in *Drosophila melanogaster*. *Dev. Biol.* **129**, 417–427 (1988).
- Ross, J. J. *et al.* *Twisted gastrulation* is a conserved extracellular BMP antagonist. *Nature* **410**, 479–483 (2001).
- Kohfeldt, E., Maurer, P., Vannahme, C. & Timpl, R. Properties of the extracellular calcium binding module of the proteoglycan tescin. *FEBS Lett.* **414**, 557–561 (1997).
- Zhu, Y., Oganessian, A., Keene, D. R. & Sandell, L. J. Type IIA procollagen containing the cysteine-rich amino propeptide is deposited in the extracellular matrix of prechondrogenic tissue and binds to TGF- β 1 and BMP-2. *J. Cell Biol.* **144**, 1069–1080 (1999).
- Pappano, W. N. *et al.* Coding sequence and expression patterns of mouse Chordin and mapping of the cognate mouse *Chrd* and human *CHRD* genes. *Genomics* **52**, 236–239 (1998).
- Lee, S., Solow-Cordero, D. E., Kessler, E., Takahara, K. & Greenspan, D. S. Transforming growth factor- β regulation of bone morphogenetic protein-1/procollagen C-proteinase and related proteins in fibrogenic cells and keratinocytes. *J. Biol. Chem.* **272**, 19059–19066 (1997).
- Cho, K. W., Blumberg, B., Steinbeisser, H. & De Robertis, E. M. Molecular nature of Spemann's organizer: the role of the *Xenopus* homeobox gene *gooseoid*. *Cell* **67**, 111–120 (1991).
- Harland, R. M. in *Methods in Cell Biology* (eds Kay, B. K. & Peng, H. B.) 685–695 (Academic, San Diego, 1991).
- Takahara, K., Lyons, G. E. & Greenspan, D. S. Bone morphogenetic protein-1 and a mammalian tolloid homologue (mTld) are encoded by alternatively spliced transcripts which are differentially expressed in some tissues. *J. Biol. Chem.* **269**, 32572–32578 (1994).
- Furuta, Y., Piston, D. W. & Hogan, B. L. M. Bone morphogenetic proteins (BMPs) as regulators of dorsal forebrain development. *Development* **124**, 2203–2212 (1997).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank M. O'Connor for the mouse Tsg cDNA clone, J. Fallon for the chick embryo cDNA library, and B. Hogan for BMP2, -4, and -7 cDNA clones. We also thank E. De Robertis and C. Niehrs for *Xenopus* chordin and Vent2 cDNAs, respectively; S. Kinoshita for technical assistance; T. Koide for technical advice on embryo bleaching; and M. O'Connor, L. Marsh and A. Hemmati-Brivanlou for communicating their work before publication. This work was supported by the NIAHS (D.S.G.), NIGMS (D.S.G. and K.W.Y.C.) and NICHD (K.W.Y.C.) of the National Institutes of Health, and by FibroGen (D.S.G.).

Correspondence and requests for materials should be addressed to D.S.G. (email: dsgrains@facstaff.wisc.edu) or K.W.Y.C. (email: kwcho@uci.edu). The GenBank accession numbers for full-length human, *Xenopus*, zebrafish and chick Tsg sequence are AF196834, AF279246, AF261692 and AF255731, respectively.

Twisted gastrulation is a conserved extracellular BMP antagonist

Jeffrey J. Ross^{*†‡}, Osamu Shimmi^{*‡}, Peter Vilmos[§], Anna Petryk^{||}, Hyon Kim^{*‡}, Karin Gaudenz[§], Spencer Hermanson^{*‡}, Stephen C. Ekker^{*‡}, Michael B. O'Connor^{*†} & J. Lawrence Marsh[§]

^{*} Departments of Genetics, Cell Biology and Development, [†] Howard Hughes Medical Institute, ^{||} Department of Pediatrics and [§] The Arnold and Mabel Beckman Center for Transposon Research, University of Minnesota, Minneapolis, Minnesota 55455, USA

[§] Department of Developmental and Cell Biology, University of California, Irvine, California 92697, USA

[‡] These authors contributed equally to this work

Bone morphogenetic protein (BMP) signalling regulates embryonic dorsal–ventral cell fate decisions in flies, frogs and fish¹. BMP activity is controlled by several secreted factors including the antagonists chordin and short gastrulation (SOG)^{2,3}. Here we show that a second secreted protein, Twisted gastrulation (Tsg)⁴, enhances the antagonistic activity of Sog/chordin. In *Drosophila*, visualization of BMP signalling using anti-phospho-Smad staining⁵ shows that the *tsg* and *sog* loss-of-function phenotypes are very similar. In S2 cells and imaginal discs, TSG and SOG together make a more effective inhibitor of BMP signalling than either of them alone. Blocking Tsg function in zebrafish with morpholino oligonucleotides causes ventralization similar to that

produced by chordin mutants. Co-injection of sub-inhibitory levels of morpholines directed against both Tsg and chordin synergistically enhances the penetrance of the ventralized phenotype. We show that Tsgs from different species are functionally equivalent, and conclude that Tsg is a conserved protein that functions with SOG/chordin to antagonize BMP signalling.

TSG is required to specify the dorsal-most structures in the *Drosophila* embryo, for example amnioserosa⁴. Mutations in the BMP-like ligands, Decapentaplegic (DPP) and Screw (SCW), the BMP inhibitory factor SOG, or the SOG-processing enzyme Tolloid (TLD), also cause loss of the amnioserosa, even though some of these products seem to have opposing biochemical functions^{2,6–8}. To place TSG activity relative to the biochemical function of these other factors, we examined its loss-of-function phenotype using molecular markers (Fig. 1A). The phenotype of *tsg* mutants (Fig. 1A, c, g, l) is most similar to that produced by loss of the BMP antagonist SOG (Fig. 1A, b, f, k) rather than that produced by loss of the ligands DPP or SCW (data not shown), or the SOG-processing protease TLD (Fig. 1A, d, h, m). In both *tsg* and *sog* mutants, the dorsal marker Rhomboid (*rho*) expands (Fig. 1A, b, c)⁹, whereas in *tld* mutants no *rho* expression is observed (Fig. 1A, d)⁸. In contrast, mutations in *tsg*, *sog* and *tld* eliminate expression of other presumptive amnioserosa markers including Race (Fig. 1A, g, f, h), Hindsight (data not shown) and Zerknullt (*zen*) (data not shown). To determine whether the response of these is indicative of different threshold levels of DPP signalling¹⁰, we used an anti-phospho-Smad antibody⁵ to directly visualize the levels of ligand signalling. Wild-type embryos accumulate phosphorylated mother against DPP (P-MAD) in an 18–20-cell-wide dorsal stripe at mid-cellulariza-

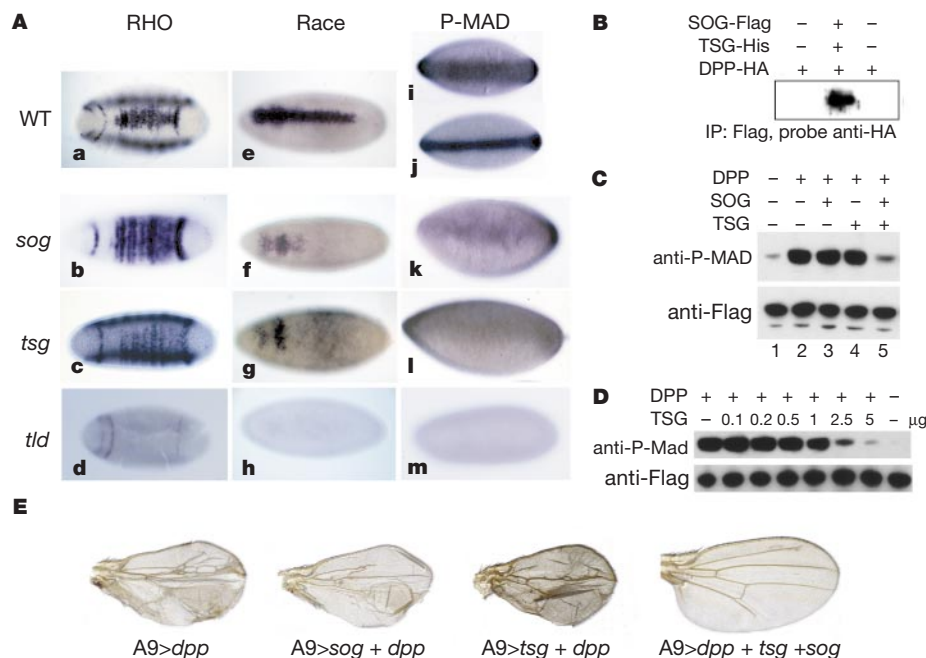


Figure 1 Tsg and SOG synergistically inhibit DPP signalling. **A**, The *tsg* loss-of-function phenotype is similar to that of *sog*. **a–h**, *In situ* hybridization of a Rhomboid RNA probe and a Race RNA probe to wild-type (WT), *sog*^{Y506}, *tsg*^{X886} and *tld*^{P4} mutant embryos. Mutant embryos were identified by the absence of *lacZ* expression supplied by a marked balancer chromosome. All embryos are at the mid-cellular-blastoderm or early gastrulation stage; anterior is to the left and the view is dorsal. **i–m**, Anti-phospho-MAD staining of WT (**i**, **j**), *sog*^{Y506} (**k**), *tsg*^{X886} (**l**) and *tld*^{P4} (**m**) mutant embryos. The embryos in **i** and **j** are dorsal side up, and in **k–m** they are viewed laterally. **B**, TSG and SOG form a high-affinity complex with DPP. S2 cells were transfected with DPP-HA and either SOG alone or SOG and TSG. After induction with Cu²⁺, the supernatants were harvested and immunoprecipitated (IP) with anti-Flag antibody, transferred to a PVDF membrane and

probed with anti-HA (12CA5 Roche) antibody as described⁹. **C**, TSG and SOG synergistically inhibit DPP signalling in S2 cells. In lanes 2–5, 20 ng purified DPP was added. Lane 1 was mock treated with buffer. In lanes 3–5, 1.25 μg SOG, 1 μg TSG or 1 μg of each were premixed with DPP and added to cells. The top panel was probed with anti-P-MAD antibody; the bottom with anti-Flag antibody. **D**, TSG alone inhibits DPP signalling in S2 cells at high concentrations. Cells were transfected with MAD-Flag as above, and then treated with 20 ng DPP and the indicated levels of TSG. **E**, TSG and SOG together form an effective inhibitor of DPP *in vivo*. Transformant lines containing UAS *dpp*, *sog* and *tsg* constructs^{9,30} were crossed in the combinations indicated to the GAL4-A9 driver³⁰.

tion (Fig. 1A, i) that rapidly resolves into an 8–9-cell-wide stripe (Fig. 1A, j) of more intensely stained cells just as gastrulation starts. Although an underlying gradient of activity not detectable by this method may exist^{7,11,12} these results instead suggest that DPP/SCW activity is distributed in a sharp on–off pattern that resolves into a narrow stripe of dorsal cells, which—posterior to the cephalic furrow—corresponds in width to those cells labelled by the amnioserosa markers *Race* and *Hindsight*. In *sog* and *tsg* mutants (Fig. 1A, k, l), P-MAD fails to refine and intensify, whereas in *tld* mutants (Fig. 1A, m) P-MAD activity is below the level of detection in all dorsal cells. We suggest that the low, uniform levels of P-MAD seen in *sog* and *tsg* mutants are sufficient to activate *rho*, but not *race*, *hnd* or *zen* transcription.

As the phenotypes of *tsg* and *sog* mutants are similar, we sought to determine whether TSG can enhance the binding of SOG to ligand. Co-immunoprecipitation of DPP by SOG is greatly enhanced when these two factors are coexpressed in S2 cells along with TSG (Fig. 1B). To test whether the combination of SOG and TSG blocks DPP signalling better than SOG alone, we developed an S2 cell-culture assay for DPP signalling (Fig. 1C). At high concentration TSG alone can block DPP signalling (Fig. 1D); however, at lower concentration, the combination of TSG and SOG together

dramatically reduces the DPP-dependent accumulation of P-MAD much more efficiently than either could alone. *In vivo* overexpression of *sog* and *tsg* together can completely reverse the phenotype of ectopic *dpp* expression in the wing, whereas the expression of either alone has no effect. We conclude that a complex of TSG and SOG is an efficient antagonist of DPP signalling.

To determine whether Tsg is conserved among other species, we sought and found genes in the database related to *Drosophila* TSG in human, mouse, zebrafish and *Xenopus*. In addition, we found a second *tsg*-related sequence in *Drosophila* (*tsg2*) and obtained a second zebrafish *tsg* (*tsg1*) using degenerate polymerase chain reaction (PCR) methods. The protein products show extensive similarity with about 50% of 202 amino-acid residues matching in all four species (see <http://darwin.bio.uci.edu/~marshlab/>). The pairs of *tsg* genes in fly and fish are closer to each other than to *tsg* in any other species, suggesting independent gene-duplication events in these two species. We mapped the human, mouse and zebrafish (*tsg1*) genes by a combination of fluorescence *in situ* hybridization (FISH) or radiation hybrid mapping. The mouse gene maps to 17E1.3–E2, a region that is syntenic to 18p11.2–3 where the human homologue resides. In zebrafish, *tsg1* is located at linkage group 24–74.5, which is syntenic to the human locus and indicates that all

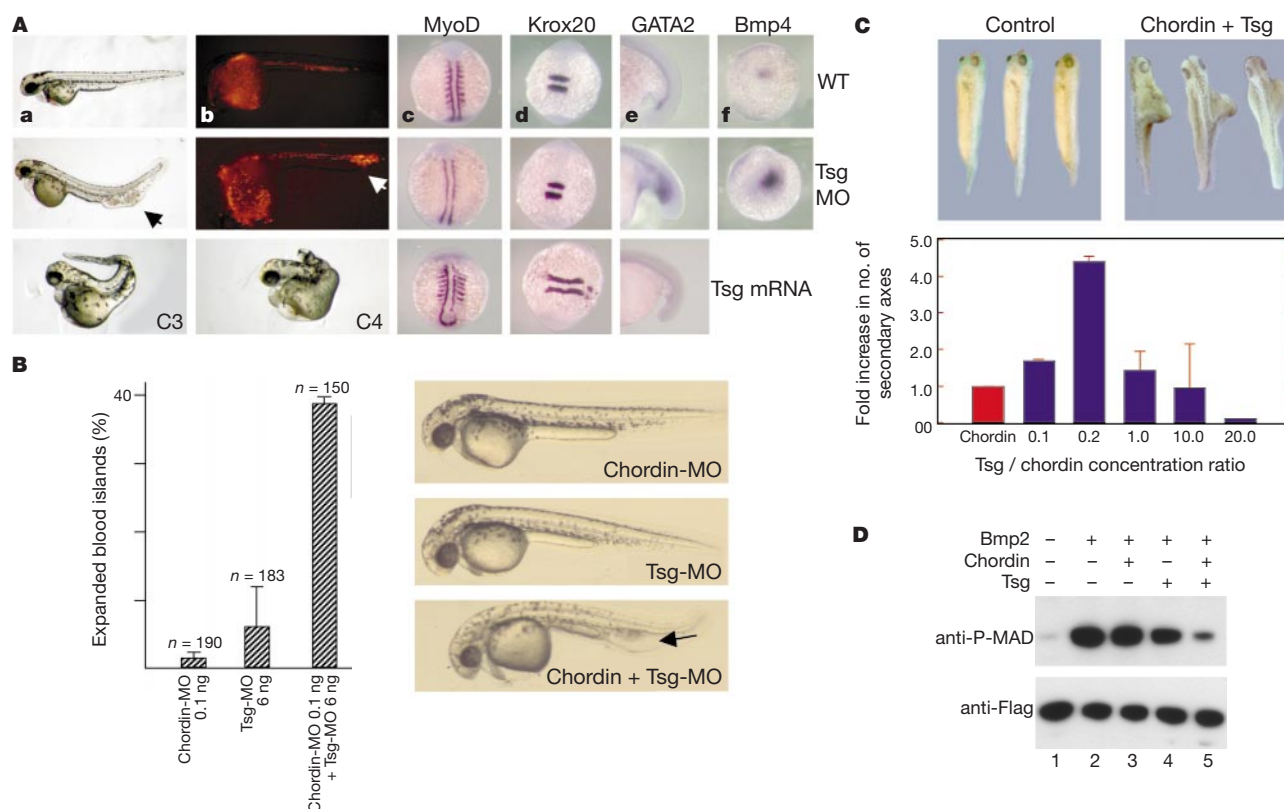


Figure 2 Vertebrate Tsg enhances chordin's antagonist function. **A**, Loss of zebrafish *tsg1* activity ventralizes zebrafish embryos whereas ectopic *tsg1* mRNA has dorsalizing activity. **a**, WT embryo (48 h). **b**, WT embryo injected with 9 ng UroD morpholino oligonucleotides. Blood cells are indicated by red staining. **c–f**, *In situ* hybridization. *myoD* (8 somite stage) (**c**); *krox20* (8 somite stage) (**d**); *GATA2* (22 somite stage) (**e**); and *bmp4* (tail bud view, 3 somite stage) (**f**). Second row as above but injected with 12 ng *tsg1* morpholino oligonucleotide. The filled and open arrows indicate ectopic blood islands (60%, $n = 110$, filled arrow; 60%, $n = 100$, open arrow). Caudal expression of *bmp4* is expanded at the 3 somite stage (48%; $n = 25$), and the blood marker *GATA2* is expanded at the 22 somite stage of development (48%; $n = 21$). Paraxial expression of *MyoD* is significantly reduced at the 8 somite stage (38%; $n = 33$), and the anterior ectodermal marker *Krox20* is moderately reduced at the 8 somite stage (49%; $n = 39$). The third row shows embryos after injection with zebrafish *Tsg1* mRNA. At 48 h of development, *Tsg1*

mRNA-injected embryos display phenotypes indistinguishable from the C3–C4 class of dorsalized mutants¹⁷. Expression of *myoD* (50%; $n = 8$) and *krox20* (70%; $n = 10$) is also significantly expanded at the 8 somite stage, while *GATA2* and *bmp4* are reduced. **B**, Enhancement of the zebrafish *Tsg1* loss-of-function phenotype by sub-inhibitory loss of the chordin gene. Embryos were injected with a low dose of zebrafish *Tsg1*-MO, chordin-MO, or both, and assayed for blood island expansion (arrow). **C**, Effect of *Xenopus* Tsg on the ability of chordin to block BMP signalling in *Xenopus*. Chordin and Tsg were injected at the ratios indicated. The graph represents the combined results from five experiments. About 200 embryos were injected for each point. On the y axis 1.0 for chordin corresponds to 2–16% induction of secondary axes. The chordin mRNA concentration was 5 pg. **D**, Vertebrate Tsg and SOG synergistically inhibit BMP-2 signalling in *Drosophila* S2 cells. The experiment was carried out as in Fig. 1. The mouse protein concentrations were: *bmp2*, 0.2 ng; chordin, 1 μ g; and Tsg, 0.5 μ g.

three genes are probably functional orthologues.

The zebrafish *tsg1* gene is expressed uniformly in early embryos, whereas zebrafish *tsg2* is only expressed at later stages (data not shown). Hence, we focused our analysis on zebrafish *tsg1* and used morpholino oligonucleotides¹³ to reduce the function of this gene in early zebrafish development. Injection of a *tsg1* morpholino oligonucleotide (*ztsg*-MO) produces a phenotype characteristic of expanded BMP signalling (Fig. 2A)^{14,15}. Using morphological criteria and fluorescent red blood cells¹³, we found that embryos develop expansions of the ventral fin region that correspond to ectopic blood islands (Fig. 2A, arrowheads), a tissue derived from ventral mesoderm. Injected embryos also show an expansion of GATA2, loss of paraxial mesoderm (visualized with the marker *myoD*), and a mild reduction of anterior ectodermal tissues (detected by staining for *krox20*). Caudal expression of *bmp4* is also expanded in these embryos (Fig. 2A), while the anterior ectodermal marker *otx2* is reduced (data not shown). Treated embryos also exhibit an expansion in apoptotic cells ventral to the yolk extension (data not shown), similar to *dino* and *mercedes* mutants^{14,16}. Overall, this phenotype is very similar to that of *ogon/mercedes* mutants^{14,15} and moderate *chordin* loss-of-function mutants, and represents a modest ventralized phenotype^{13,14}.

Increasing the level of zebrafish *tsg1* by injecting messenger RNA produced phenotypes characteristic of diminished BMP signalling^{17–19} including reduced axial length with loss of ventral fin (Fig. 2A), an expansion of *myoD* and *krox20*, and a reduction in GATA2. This is a phenocopy of the C3–C4 class of dorsalized mutant embryos, similar to that of the Snailhouse (BMP7 homologue) and Piggytail mutations^{17,19}. Furthermore, the dorsalizing effect of zebrafish Tsg1 mRNA partially reverses the ventralizing

effect of *tsg1* (9 ng *tsg1*-MO caused $47 \pm 2\%$, $n = 376$, ventralized embryos; 9 ng *tsg1*-MO plus 30 pg Tsg1 mRNA resulted in $19 \pm 8\%$, $n = 270$, ventralized embryos) suggesting that loss of *tsg1* is responsible for the phenotype. We conclude that loss of *tsg1* leads to embryos with a ventralized phenotype, whereas ectopic expression of *tsg1* leads to a dorsalized embryonic phenotype.

As our *Drosophila* data suggested that one function of TSG is to co-operate with SOG to inhibit BMP signalling, we asked whether the same relationship is true in vertebrates by determining whether a modest reduction of zebrafish *chordin* activity could enhance the effect of a moderate reduction in *tsg1* activity. Sub-inhibitory levels of a zebrafish *chordin* morpholino oligonucleotide and *tsg1*-MO were injected into wild-type embryos, and the effect on ectopic blood island development was scored. These two morpholino oligonucleotides synergistically enhanced blood island expansion (Fig. 2B), supporting the view that both of these gene products co-operatively inhibit BMP signalling. As with the *Drosophila* components, we found that the combination of purified mouse chordin and Tsg was better able to inhibit mouse BMP-stimulated phosphorylation of Mad in S2 cells than either could alone (Fig. 2D).

We also tested for synergy between Tsg and chordin mRNA in *Xenopus* embryos by co-injecting their mRNAs and scoring for enhancement of secondary axis formation²⁰. Co-injection of *Xenopus* Tsg and chordin reveals a dose–response optimum. When a sub-inhibitory dose of chordin mRNA is supplemented with increasing levels of Tsg mRNA, the fraction of embryos exhibiting a secondary axis increases up to 4.5-fold over chordin alone at a 1/5 ratio of Tsg/chordin mRNA. However, if the Tsg/chordin ratio is increased to 1:1 or higher, the number of secondary axes is reduced to basal levels and the resulting tadpoles have normal morphology. Injection of 150 pg Tsg alone (the highest concentration of Tsg mRNA used in these experiments) had no effect on embryonic development. Notably, if we increase the level of Tsg relative to chordin in the S2 experiments, we do not see a reversal of the inhibition phenotype (data not shown), suggesting that additional factors probably modulate the *in vivo* response. Taken together, we conclude that, like *Drosophila* TSG, vertebrate Tsg can co-operate with chordin to inhibit BMP signalling.

As a final test of the functional equivalence of the vertebrate and invertebrate *tsg* genes, we expressed the human and mouse genes under the control of the UAS promoter in flies, and injected *Drosophila* TSG mRNA into zebrafish embryos. The phenotype of animals expressing human TSG and *Drosophila sog* in wing discs (Fig. 3a) resembles that of *dpp* shortvein alleles²¹ and is very similar to that produced by coexpression of the *Drosophila tsg* and *sog* genes (Fig. 3a; see also ref. 9). When injected into zebrafish, *Drosophila tsg* produces a dorsalized phenotype equivalent to that produced by zebrafish *tsg1*, which includes reduced axial length and expansion of *krox20* (Fig. 2A; compare with Fig. 3b) and *myoD* (data not shown).

Our experiments, and those of others^{22–24}, suggest that Tsg has three molecular functions. First, it can synergistically inhibit Dpp/BMP action in both *Drosophila* and vertebrates by forming a tripartite complex between itself, SOG/chordin and a BMP ligand (Fig. 1B, see also refs 9, 24). Second, Tsg seems to enhance the Tld/BMP-1-mediated cleavage rate of SOG/chordin and may change the preference of site utilization (O.S. and M.B.O., unpublished observations; see also refs 9, 23). Third, Tsg can promote the dissociation of chordin cysteine-rich (CR)-containing fragments from the ligand²⁴. Different organisms may exploit each of these properties to different degrees during development depending on the relative *in vivo* concentrations of each molecule. We propose that in *Drosophila* and zebrafish the primary function of Tsg is to form a tripartite complex between itself, Sog/chordin and a BMP ligand. In *Drosophila*, this complex acts to redistribute a limiting amount of DPP, such that activity is elevated dorsally at the expense of being lowered laterally. The net driving force for this redistribution is likely to be diffusion of SOG from its ventral source of synthesis²⁵.

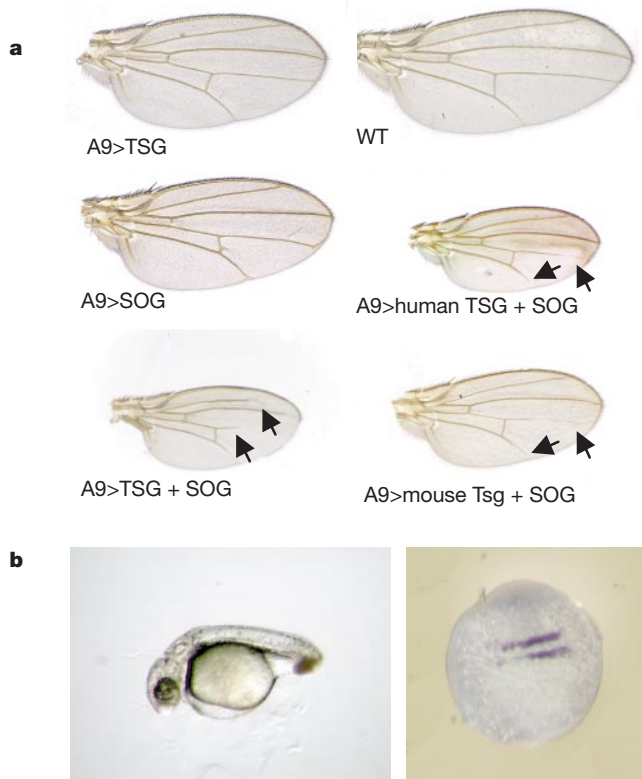


Figure 3 Functional equivalence of Tsg proteins. **a**, The GAL4-A9 driver was crossed to flies at 25 °C (first column) or 29 °C (second column) with the indicated transgenes. Arrows indicate loss of distal vein material. **b**, Injection of 30 pg *Drosophila* Tsg into zebrafish embryos results in a C3-like dorsalized phenotype (54%; $n = 22$) at 48 h after fertilization (left). Right, embryo injected with 30 pg *Drosophila tsg* exhibiting expansion of *krox20* at the 8 somite stage (compare with Fig. 2a).

This is consistent with the finding that SOG diffusion is essential for activation of genes such as *race* that require high levels of DPP/SCW signalling (Fig. 1, see also ref. 26). In this model TLD would serve to modulate both the net movement of DPP and its release from the inhibitory complex by cleaving SOG^{8,25}. The ability of TSG to enhance the rate of SOG cleavage may also be an important aspect of this model in that it helps ensure the proper timing of these rapid developmental events. It seems unlikely that TSG is needed to remove an inhibitory CR-containing fragment from DPP as the affinity of full-length SOG for DPP in the absence of TSG seems to be low. Likewise, in zebrafish the phenotype of reduced Tsg function is ventralized and not dorsalized as would be predicted if Tsg were primarily needed to release inhibitory CR fragments from ligand. In *Xenopus*, however, perhaps the endogenous levels of full-length chordin and CR fragments are higher than in zebrafish, thereby making the CR displacement activity of Tsg the more important biological function²⁴. Determination of the *in vivo* levels of these proteins, along with a more careful analysis of the concentration optima for each type of reaction involving Tsg function, will be required before we can fully understand all of its *in vivo* activities. □

Methods

Isolation of *tsg* clones and gene mapping

Human TSG complementary DNA clones (accession numbers AW160804, AA905905, AI222228, AA486291, AI018381, AI379897 and AA758784) were obtained from Research Genetics. One clone (AI018381) was sequenced in its entirety, additional 5' sequence was obtained from published Est sequences. Mouse Tsg cDNA clone (accession number AW258143) was also obtained from Research Genetics. The zebrafish *tsg1* was isolated from an epiboly cDNA library (S. Ekker) using two degenerate primers (5' primer: 5'-TG(CT)TG(CT)AA(AG)GA(CTAG)TG(CT)(CTA)T-3' and 3' primer: 5'-CC(CTG)A(CT)(AG)GA(CT)TC(AG)CA(AG)CA-3'). These primers amplified a 0.5-kilobase (kb) fragment that was used as a probe to identify a 1.2 kb cDNA from a zebrafish epiboly library. We sequenced this clone using standard methods. The human TSG locus was mapped against the Stanford G3 hamster-human radiation hybrid panel using primer pairs at the beginning, middle and end of the TSG mRNA. This placed human TSG between STS markers D18, and D18 within cytogenetic band 18p11.2. The mouse Tsg was mapped by FISH using a 16-kb genomic mouse Tsg fragment as a probe. The chromosomal assignment and band designation were determined by sequential G-banding to FISH. The zebrafish (*Danio rerio*) *tsg1* gene was mapped to linkage group 24 at 74.5 cM using a mouse-fish radiation hybrid panel²⁷.

Altering signal peptides

While conducting these studies, we found that the secretion signals of the mammalian and *Drosophila* genes are incompatible with the other species. To circumvent these secretion-related problems, we used PCR to replace the human and mouse signal sequences with the *Drosophila* sequence and also the *Drosophila* signal peptide with the zebrafish sequence (details are available on request).

Production and purification of recombinant proteins

Recombinant proteins SOG-Myc, Tsg-His and Dpp-haemagglutinin (HA) were produced as described⁹. Conditioned medium containing SOG-Myc was applied to a 1 × 10-cm² S-Sepharose column (Pharmacia) equilibrated with 100 mM MOPS-Na, pH 6.0 (buffer A). After washing with buffer A containing 300 mM NaCl, the column was eluted with buffer A containing 750 mM NaCl, and the fractions were combined and stored for further use. Conditioned medium containing Tsg-His was applied to a 1 × 10-cm Q-Sepharose column (Pharmacia) equilibrated with 50 mM Tris-HCl, pH 7.5 (buffer B). After washing with buffer B containing 200 mM NaCl, the column was eluted with buffer B containing 500 mM NaCl, and the fractions were combined and applied to a 1 × 4-cm Ni-NTA agarose column (QIAGEN) equilibrated with 100 mM Tris-HCl, pH 8.0 (buffer C). After washing with buffer C containing 1 M NaCl, the column was eluted with buffer C containing 100 mM imidazole. Fractions containing Tsg-His were pooled and dialysed against 50 mM Tris-HCl, 150 mM NaCl, pH 7.4.

Signalling assays

Ten micrograms of Flag-tagged MAD were transfected in S2 cells at 2 × 10⁷ cells per dish. After 3 days, the cells were collected and split into 20 samples. One microgram Tsg-His and/or 1.25 µg SOG-Myc (Fig. 1C), or 0.5 µg mouse Tsg-protein C and/or 1 µg chordin-His (R&D Systems) (Fig. 2D) were premixed for 3 h at room temperature (RT) with 10⁻⁹ M Dpp or 10⁻¹¹ M BMP2 (R&D Systems) and then incubated with S2 cells expressing Flag-Mad for 3 h at RT. The cells were spun down and lysed by 1 × SDS-PAGE buffer. The supernatants were separated by SDS-PAGE and transferred to polyvinylidene difluoride (PVDF) membranes (Millipore). The membrane was probed with anti-Phospho Mad PS1 antibody at 1/5,000 dilution⁵ and anti-Flag M2 antibody (Kodak) at 1/2,000 dilution,

followed by incubation in secondary antibody (horseradish peroxidase-conjugated goat anti-rabbit or anti-mouse; Jackson Laboratory) and developed using ECL substrate (Pierce).

Morpholino oligonucleotides

We obtained Morpholino oligonucleotides from Gene Tools, LLC (Corvallis). We selected sequences on the basis of the design parameters recommended by the company. The zebrafish chordin morpholino oligonucleotide was as described¹³. *tsg1*-MO 5'-CTGATG ATGATGATGAAGACCCCAT-3'.

Embryo manipulations and microinjections

Morpholino oligonucleotides were injected as described¹³. For *Xenopus* injections, embryos were obtained by *in vitro* fertilization and cultured as described²⁸. Microinjections of mRNA were performed at the 4-cell stage in 0.3 × MMR, 3.5% Ficoll. We determined dorsal-ventral polarity of early cleavage stage embryos using pigmentation differences²⁸.

In situ hybridization and antibody staining

Hybridization to *Drosophila* and zebrafish embryos was as described^{4,29}. The rabbit anti-phospho Mad antibody was a gift from P. ten Dijke and used at 1/2,000 dilution. Staining was visualized using an alkaline phosphatase-coupled secondary antibody (Promega Laboratories).

Received 29 August 2000; accepted 8 January 2001.

- Holley, S. A. & Ferguson, E. L. Fish are like flies are like frogs: conservation of dorsal-ventral patterning mechanisms. *BioEssays* **19**, 281-284 (1997).
- Ferguson, E. L. & Anderson, K. V. Localized, enhancement and repression of the activity of the TGF-β family member, *decapentaplegic*, is necessary for dorsal-ventral pattern formation in the *Drosophila* embryo. *Development* **114**, 583-597 (1992).
- Piccolo, S., Sasaki, Y., Lu, B. & De Robertis, E. M. Dorsal-ventral patterning in *Xenopus*: inhibition of ventral signals by direct binding of Chd to BMP-4. *Cell* **86**, 589-598 (1996).
- Mason, E. D., Konrad, K. D., Webb, C. D. & Marsh, J. L. Dorsal midline fate in *Drosophila* embryos requires twisted gastrulation, a gene encoding a secreted protein related to human connective tissue growth factor. *Genes Dev.* **8**, 1489-1501 (1994).
- Tanimoto, H., Itoh, S., ten Dijke, P. & Tabata, T. Hedgehog creates a gradient of DPP activity in *Drosophila* wing imaginal discs. *Mol. Cell* **5**, 59-71 (2000).
- Ray, R. P., Arora, K., Nüsslein-Volhard, C. & Gelbart, W. M. The control of cell fate along the dorsal-ventral axis of the *Drosophila* embryo. *Development* **113**, 35-54 (1991).
- Arora, K. & Nüsslein-Volhard, C. Altered mitotic domains reveal fate map changes in *Drosophila* embryos mutant for zygotic dorsoventral patterning genes. *Development* **114**, 1003-1024 (1992).
- Marques, G. et al. Production of a DPP activity gradient in the early *Drosophila* embryo through the opposing actions of the SOG and TLD proteins. *Cell* **91**, 417-26 (1997).
- Yu, K. et al. Processing of the *Drosophila* Sog protein creates a novel BMP inhibitory activity. *Development* **127**, 2143-2154 (2000).
- Ashe, H. L., Mannervik, M. & Levine, M. Dpp signaling thresholds in the dorsal ectoderm of the *Drosophila* embryo. *Development* **127**, 3305-3312 (2000).
- Ferguson, E. L. & Anderson, K. V. Decapentaplegic acts as a morphogen to organize dorsal-ventral pattern in the *Drosophila* embryo. *Cell* **71**, 451-461 (1992).
- Wharton, K. A., Ray, R. P. & Gelbart, W. M. An activity gradient of *decapentaplegic* is necessary for the specification of dorsal pattern elements in the *Drosophila* embryo. *Development* **117**, 807-822 (1993).
- Nasevicius, A. & Ekker, S. C. Effective targeted gene 'knockdown' in zebrafish. *Nature Genet.* **26**, 216-220 (2000).
- Hammerschmidt, M. et al. *dino* and *mercedes*, two genes regulating dorsal development in the zebrafish embryo. *Development* **123**, 95-102 (1996).
- Miller-Bertoglio, V. et al. Maternal and zygotic activity of the zebrafish *ogon* locus antagonizes BMP signaling. *Dev. Biol.* **214**, 72-86 (1999).
- Fisher, S., Amacher, S. L. & Halpern, M. E. Loss of cerebium function ventralizes the zebrafish embryo. *Development* **124**, 1301-1311 (1997).
- Mullins, M. C. et al. Genes establishing dorsoventral pattern formation in the zebrafish embryo: the ventral specifying genes. *Development* **123**, 81-93 (1996).
- Kishimoto, Y., Lee, K. H., Zon, L., Hammerschmidt, M. & Schulte-Merker, S. The molecular nature of zebrafish swirl: BMP2 function is essential during early dorsoventral patterning. *Development* **124**, 4457-4466 (1997).
- Dick, A. et al. Essential role of Bmp7 (snailhouse) and its prodomain in dorsoventral patterning of the zebrafish embryo. *Development* **127**, 343-354 (2000).
- Harland, R. & Gerhart, J. Formation and function of Spemann's organizer. *Annu. Rev. Cell Dev. Biol.* **13**, 611-667 (1997).
- Segal, D. & Gelbart, W. M. Shortvein, a new component of the decapentaplegic gene complex in *Drosophila melanogaster*. *Genetics* **109**, 119-143 (1985).
- Chang, C. et al. Twisted gastrulation can function as a BMP antagonist. *Nature* **410**, 483-487 (2001).
- Scott, I. C. et al. Homologues of Twisted gastrulation are extracellular cofactors in antagonism of BMP signalling. *Nature* **410**, 475-476 (2001).
- Oelgeschlager, M., Larrain, J., Geisler, D. & De Robertis, E. M. The evolutionarily conserved BMP-binding protein Twisted gastrulation promotes BMP signalling. *Nature* **405**, 757-763 (2000).
- Holley, S. A. et al. The *Xenopus* dorsalizing factor noggin ventralizes *Drosophila* embryos by preventing DPP from activating its receptor. *Cell* **86**, 607-617 (1996).
- Ashe, H. L. & Levine, M. Local inhibition and long-range enhancement of Dpp signal transduction by Sog. *Nature* **398**, 427-431 (1999).
- Hukriede, N. A. et al. Radiation hybrid mapping of the zebrafish genome. *Proc. Natl. Acad. Sci. USA* **96**, 9745-9750 (1999).

28. Cho, K. W., Blumberg, B., Steinbeisser, H. & De Robertis, E. M. Molecular nature of Spemann's organizer: the role of the *Xenopus* homeobox gene gooseoid. *Cell* **67**, 1111–1120 (1991).
29. Jowett, T. Analysis of protein and gene expression. *Methods Cell Biol.* **59**, 63–85 (1999).
30. Haerry, T. E., Khalsa, O., O'Connor, M. B. & Wharton, K. A. Synergistic signaling by two BMP ligands through the SAX and TKV receptors controls wing growth and patterning in *Drosophila*. *Development* **125**, 3977–3987 (1998).

Acknowledgements

We are grateful to K. Cho, D. Greenspan and A. Hemmati-Brivanlou for communication of results before publication and to E. De Robertis for comments on the manuscript. We thank M. Tsang at R&D systems for a gift of purified mouse chordin; D. Greenspan for purified mouse Tsg-C protein; J. Groppe for purified Dpp; P. ten Dijke for the gift of anti-phospho Mad; and E. De Robertis for the *Xenopus* Tsg cDNA clone. We also thank E. Hirsch for assistance with the mouse FISH analysis. This work was supported by NIH grants to J.L.M., M.B.O. and S.C.E. O.S. was supported by Nippon Roche K.K. A.P. was supported by a training grant from the NIH. M.B.O. is an Associate Investigator for the Howard Hughes Medical Institute.

Correspondence and requests for materials should be addressed to M.B.O. (e-mail: moconnor@mail.med.umn.edu) or J.L.M. (e-mail: jlmarsh@uci.edu). Zebrafish *tsg1* is deposited in GenBank under accession number AF332096.

Twisted gastrulation can function as a BMP antagonist

Chenbei Chang*, Douglas A. Holtzman†‡, Samantha Chau†, Troy Chickering†, Elizabeth A. Woolf†, Lisa M. Holmgren†, Jana Bodorova†‡, David P. Gearing†‡, William E. Holmes† & Ali H. Brivanlou*

* Laboratory of Vertebrate Molecular Embryology, The Rockefeller University, Box 32, 1230 York Avenue, New York, New York 10021, USA

† Millennium Pharmaceuticals, 620 Memorial Drive, Cambridge, Massachusetts 02139, USA

Bone morphogenetic proteins (BMPs), including the fly homologue Decapentaplegic (DPP), are important regulators of early vertebrate and invertebrate dorsal–ventral development^{1–6}. An evolutionarily conserved BMP regulatory mechanism operates from fly to fish, frog and mouse to control the dorsal–ventral axis determination. Several secreted factors, including the BMP antagonist chordin/Short gastrulation (SOG)^{7–12}, modulate the activity of BMPs. In *Drosophila*, Twisted gastrulation (TSG) is also involved in dorsal–ventral patterning^{13–15}, yet the mechanism of its function is unclear. Here we report the characterization of the vertebrate Tsg homologues. We show that Tsg can block BMP function in *Xenopus* embryonic explants and inhibits several ventral markers in whole-frog embryos. Tsg binds directly to BMPs and forms a ternary complex with chordin and BMPs. Coexpression of Tsg with chordin leads to a more efficient inhibition of the BMP activity in ectodermal explants. Unlike other known BMP antagonists, however, Tsg also reduces several anterior markers at late developmental stages. Our data suggest that Tsg can function as a BMP inhibitor in *Xenopus*; furthermore, Tsg may have additional functions during frog embryogenesis.

We isolated human Twisted gastrulation (TSG) in a screen for secreted factors, and mouse and *Xenopus* Tsg by low-stringency hybridization using human TSG as the probe. These vertebrate Tsgs have a high sequence homology to each other (more than 80% identical) and are about 30% identical to *Drosophila* TSG at the amino-acid level (data not shown). Tsg is expressed maternally and

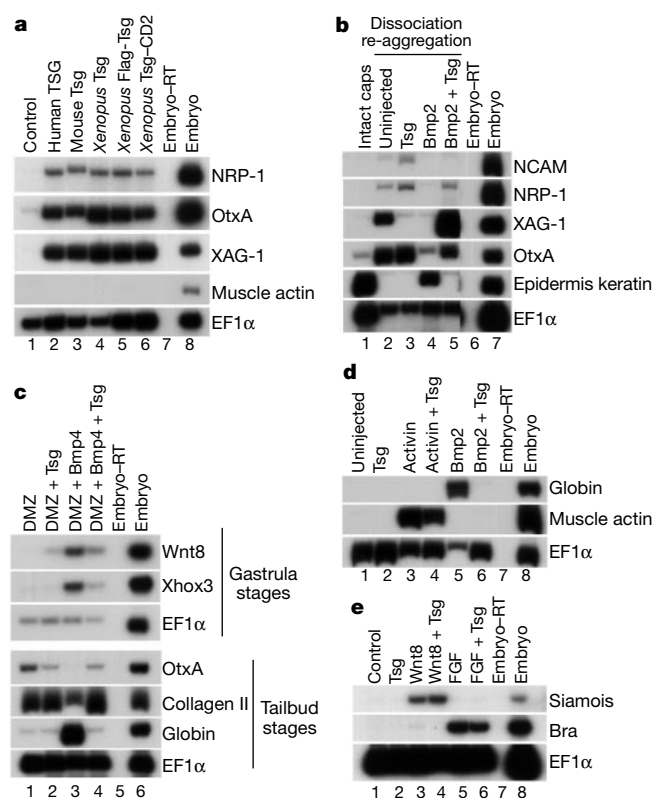


Figure 1 Vertebrate Tsgs inhibit BMP signalling in embryonic explants. **a**, Vertebrate Tsgs induce expression of cement gland and neural markers in animal caps. **b**, *Xenopus* Tsg blocks epidermal induction and neural inhibition by Bmp2 in dissociated animal-cap cells. **c**, *Xenopus* Tsg blocks ventralization of dorsal marginal-zone explants by Bmp4. **d**, *Xenopus* Tsg blocks mesodermal induction by Bmp2, but not activin. **e**, *Xenopus* Tsg does not interfere with the Wnt or FGF signalling. In animal-cap assays, RNAs were injected into animal poles of both cells of 2-cell-stage embryos. Animal caps were dissected at blastula stages (stage 9) and incubated to gastrula (stage 11, **e**) or neurula stages (stage 20, **a**, **b**, **d**). In **b**, the caps were dissociated at blastula stages for 4 h before re-aggregation and incubation to neurula stages, as described¹⁶. In the marginal-zone assay, RNAs were injected into the two dorsal blastomeres of 4-cell-stage embryos. Dorsal marginal-zone explants were dissected at early gastrula stage (stage 10) and incubated to mid-gastrula (stage 11) or tailbud (stage 28) stages. Xhox3, Wnt8 and globin are ventral markers, whereas OtxA and collagen II are dorsal markers. The weak induction of Wnt8 is not always observed. In **e**, the basic FGF protein (Sigma) was added at 100 ng ml⁻¹ to the animal caps at the blastula stages. The amount of RNA injected into the embryos was: 2 ng, all Tsg; 0.5 ng, Bmp2; 0.5 ng, Bmp4; 5 pg, activin; and 50 pg, Wnt8.

in all developmental stages in *Xenopus*, and at least from gastrula stages onward in mouse (data not shown). Expression of Tsg is also detected in a variety of adult tissues in both mouse and human (data not shown).

To study the function of Tsgs, we first analysed their activities in *Xenopus* ectodermal explants (animal caps). As shown in Fig. 1a, human, mouse and *Xenopus* Tsg induce the cement gland and the neural markers XAG-1, OtxA and NRP-1 with comparable efficiency, suggesting that these vertebrate Tsgs function similarly in *Xenopus*. The induction of cement gland and neural markers in animal caps in the absence of mesoderm is normally associated with inhibition of the BMP signalling^{16–18}, so we therefore addressed whether Tsg could directly block the activity of BMP. We first examined the effect of Tsg on ventralization of the ectodermal cells by BMPs. As described previously¹⁶, intact animal caps express high levels of epidermal keratin. This expression is suppressed when caps from blastula stages are dissociated for 4 h (Fig. 1b, lanes 1 and

‡ Present addresses: Microbia, Inc., One Kendall Square, Building 1400W Suite 1418, Cambridge, Massachusetts 02139, USA (D.A.H.); SAIC/NCI-FCRDC, Building 560/21–50, Frederick, Maryland 21702, USA (J.B.); CSL Limited, 45 Poplar Road, Parkville, Vic 3052 Australia (D.P.G.).

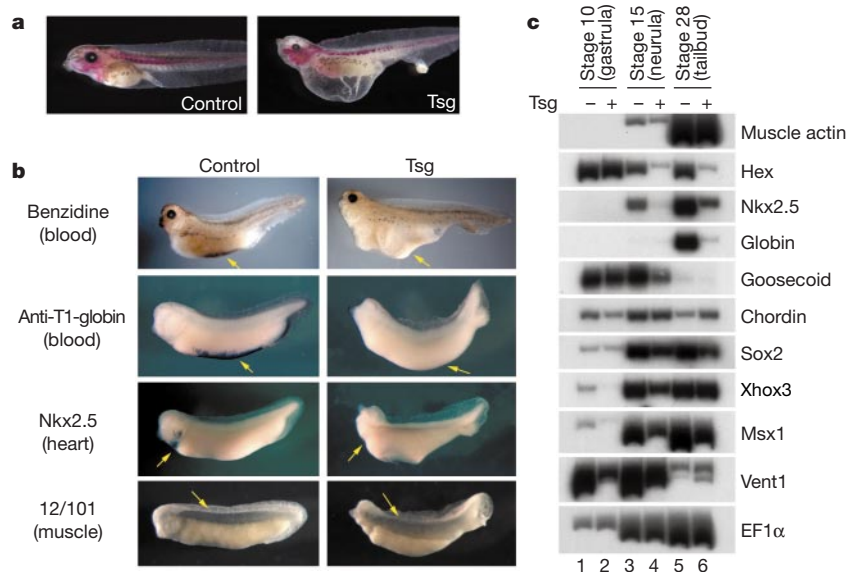


Figure 2 Overexpression of Tsg perturbs early *Xenopus* development and interferes with formation of tissues that require active BMP signalling. **a**, Tsg induces embryonic defects in tadpoles in a non-cell-autonomous manner. **b**, Tsg blocks expression of the blood (globin) and the heart (Nkx2.5) markers, but does not reduce the muscle marker. **c**, Overexpression of Tsg perturbs gene expression pattern during early *Xenopus* embryogenesis. In **a**, 0.1 ng β -Gal RNA was co-injected with 2 ng *Xenopus* Tsg RNA into one dorsal vegetal cell of 8-cell-stage embryos. The tadpole embryos were stained with

the red gal substrate. In **b** and **c**, 1 ng *Xenopus* Tsg RNA was injected into the marginal zones of all cells of the 4-cell-stage embryos. Benzidine staining was performed as described¹⁹. Whole-mount antibody staining with the muscle-specific 12/101 antibody was performed to detect the muscle formation, whereas *in situ* hybridization with globin and Nkx2.5 probes was done to analyse the blood and the heart formation, respectively.

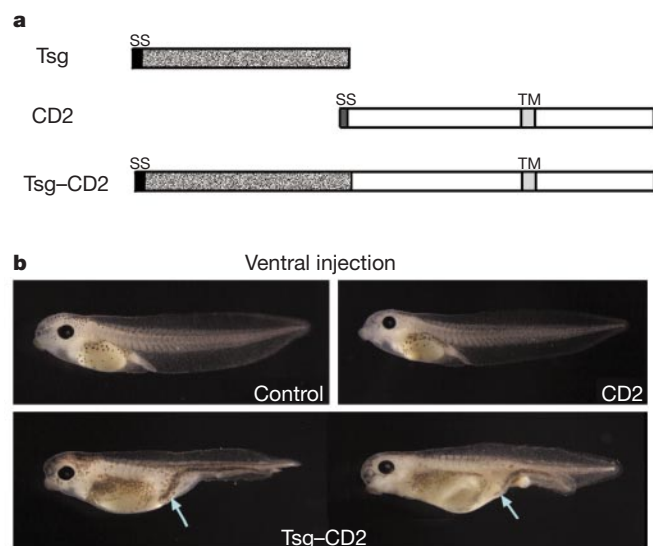


Figure 3 Membrane-tethered *Xenopus* Tsg induces partial secondary axis when expressed in the ventral region. **a**, Schematic presentation of the membrane-tethered Tsg-CD2 construct. **b**, Ventral injection of Tsg-CD2 (2 ng) induces a partial secondary axis (blue arrows). SS, signal sequence; TM, transmembrane domain.

2). Cement gland and neural markers are turned on in these dissociated samples. While Bmp2 restores the transcription of epidermal keratin and inhibits the expression of neural markers, it cannot do so in the presence of Tsg (Fig. 1b, lanes 4 and 5). These results indicate that Tsg directly antagonizes the neural inhibition and epidermal induction activity of BMPs in dissociated animal caps.

We next examined the effect of Tsg on ventralization of the mesodermal explants by BMPs. Bmp4 inhibits dorsal and induces

ventral gene expression in dorsal marginal zone (DMZ) explants (Fig. 1c, lane 3). Coexpression of *Xenopus* Tsg with Bmp4 re-establishes dorsal marker expression and reduces the ventral gene induction in these explants (Fig. 1c, lane 4), indicating that it also inhibits BMP activity in the mesodermal cells. Notably, we also detected a reduction in the level of OtxA (a marker for anterior neural tissue at tailbud stages) when Tsg alone is expressed in the DMZ (Fig. 1c, lane 2). This phenotype has not been observed with other BMP antagonists, and suggests that Tsg may be involved in processes other than inhibition of BMP signalling. To determine whether Tsg also blocks other signal transduction pathways, we assayed for the effect of *Xenopus* Tsg on marker induction by activin, basic fibroblast growth factor (FGF) and Wnt8. Tsg specifically inhibits BMP-dependent mesoderm induction, but does not interfere with the expression of the markers induced by the other signalling molecules (Fig. 1d, e). These results demonstrate that Tsg is specific for a BMP pathway and does not block activin, FGF or Wnt signalling.

We next looked at the *in vivo* function of Tsg by gain-of-function studies. Injection of Tsg RNAs from different vertebrates in early embryos induces a similar phenotype (not shown), indicating that they have similar activities *in vivo*. The embryos injected with Tsg RNA exhibit several developmental defects at tadpole stages, such as an enlargement of the dorsal fin, malformation of the proctodeum, and a reduction of the head (Fig. 2a). Lineage tracing with nuclear β -gal shows that the effect induced by Tsg is non-cell autonomous (Fig. 2a). The phenotype of Tsg-expressing embryos is unique and does not resemble the phenotypes induced by overexpression of other BMP antagonists. We therefore addressed whether Tsg affects the cells and tissues that rely on the BMP signalling. We first examined the formation of the blood cells, which are derived from ventral tissues and require active BMP signals¹⁹. In control embryos, the blood cells stained with benzidine are localized at the ventral side of the embryos; however, this staining is not observed in embryos injected with *Xenopus* Tsg RNA (Fig. 2b). Furthermore, *in situ* hybridization with a blood-specific anti-T1-globin probe

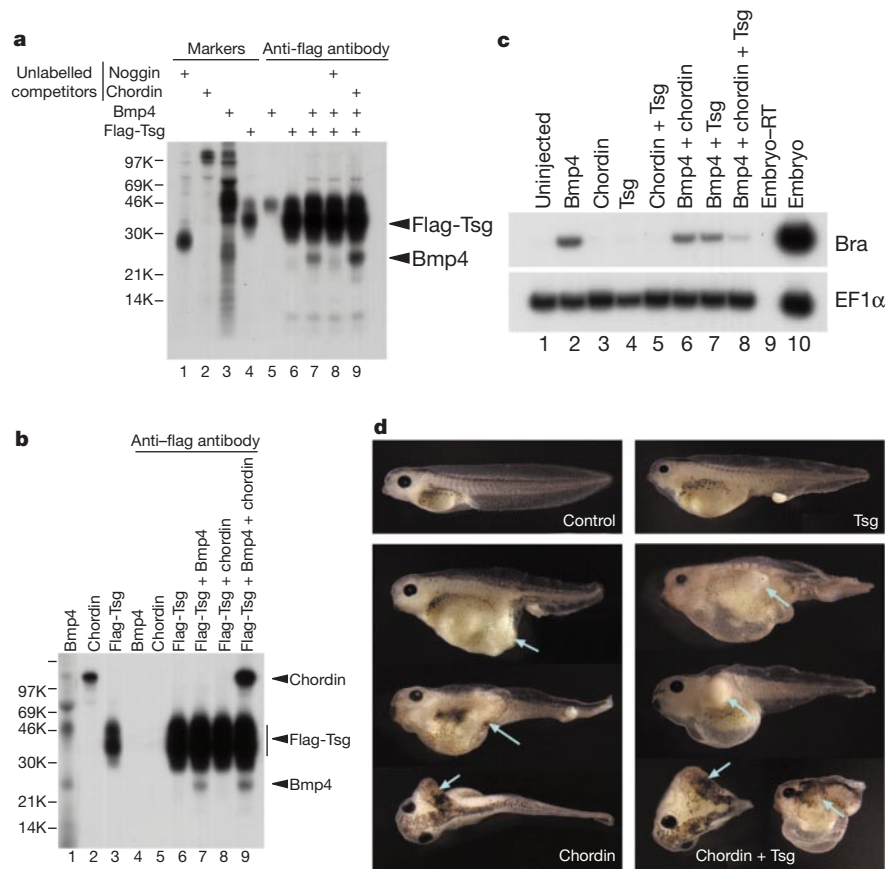


Figure 4 *Xenopus* Tsg binds to Bmp4, forms a ternary complex with chordin and Bmp4, and coexpression of chordin and Tsg leads to a more efficient inhibition of BMP activity in ectodermal explants. **a**, Tsg binds directly to Bmp4, and the binding is competed by the BMP antagonist noggin but not by chordin. **b**, Tsg forms a ternary complex with Bmp4 and chordin. **c**, Coexpression of Tsg and chordin leads to a more efficient inhibition of the BMP activity. **d**, Secondary axis induction by chordin with low Tsg/chordin ratios. The protein-binding assay was performed as described³⁰. The use of chick rather than *Xenopus*

chordin and co-immunoprecipitation in the absence of chemical crosslinkers may account for the lack of direct binding of Tsg to chordin²⁶ as reported. In **c**, 1 ng Bmp4, 0.2 ng chordin and 1 ng *Xenopus* Tsg RNAs were injected into animal poles of 2-cell-stage embryos. In **d**, 1 ng chordin and 0.2–2 ng *Xenopus* Tsg RNAs were injected into one ventral vegetal blastomere of the 8-cell-stage embryos. In **a** and **b**, the relative molecular masses of protein markers are labelled on the left side of the gels.

reveals that blood cells are absent in the Tsg-expressing embryos (Fig. 2b). We next examined the formation of the heart, which has been reported in both the chick and the frog to require BMP signalling^{20,21}. *In situ* hybridization with a heart marker, *Nkx2.5*, shows that the Tsg-expressing embryos have reduced transcription of *Nkx2.5* (Fig. 2b). In contrast, the expression of a dorsal mesoderm marker in muscle is not decreased by Tsg (Fig. 2b). Our data show that Tsg interferes with the development of several tissues that require BMP signalling.

To further analyse the effects of Tsg, we assayed for expression of both dorsal and ventral genes in Tsg-injected embryos at different developmental stages. In gastrula embryos, several ventral markers, such as *Xhox3*, *Vent1* and *Msx1*, are downregulated by *Xenopus* Tsg, whereas dorsal markers, including *Sox2*, *goosecoid* and *Hex*, are not much affected (Fig. 2c). The expression of the dorsal gene *chordin* also remains the same in most cases, although we occasionally observed a slight reduction of its level. The pattern of gene expression changes during development, so that at later stages, *Xhox3* and *Msx1* are also expressed in the neural tissue and neural crest cells. The reduction of their expression is less profound, and at tailbud stages *Xhox3* expression is normal. In agreement with our whole-mount staining results, the transcripts of both globin and *Nkx2.5* are reduced by Tsg, whereas the muscle actin remains intact at tailbud stages (Fig. 2c). Unexpectedly, we observed the reduction of several anterior genes, such as *Hex* (a marker for anterior endo-

derm) and *goosecoid* (a prechordal mesoderm gene), by *Xenopus* Tsg from neurula stages onward. *Sox2*, whose expression domain includes the anterior neural region, is also slightly downregulated. Our data demonstrate that Tsg reduces BMP signalling at early developmental stages; furthermore, Tsg may have one or more additional functions in anterior embryonic development at a stage after the onset of gastrulation.

Although Tsg can inhibit BMP function, it does not induce a partial secondary axis when injected into the ventral side of early *Xenopus* embryos. One possible explanation for this is that Tsg may be a highly diffusible molecule, as has been observed in *Drosophila*¹⁵, and does not accumulate to high concentrations at the ventral side to establish an organizer-like activity. To test this hypothesis, we constructed a membrane-tethered form of Tsg by fusing the full-length *Xenopus* Tsg in frame with an integral membrane protein CD2 (Fig. 3a). The chimaeric protein, Tsg-CD2, induces cement gland and neural markers in animal caps to the same extent as wild-type *Xenopus* Tsg (Fig. 1a), suggesting that the two proteins have similar activities. Injection of the CD2 RNA into early frog embryos does not induce any phenotype (Fig. 3b), but ventral injection of Tsg-CD2 RNA leads to embryos with partial secondary axes (88%, $n = 59$; Fig. 3b, blue arrow). These results indicate that localized Tsg behaves like other BMP antagonists and can induce ectopic dorsal structures on the ventral side of frog embryos.

To understand the mechanism of BMP inhibition by Tsg, we

performed biochemical analysis on protein interaction, using radio-labelled proteins from conditioned oocyte medium and a Flag-tagged the wild-type Tsg. *Xenopus* Flag-Tsg, which has similar activities as the wild-type Tsg (Fig. 1a; and data not shown), binds to Bmp4 directly, as determined by the co-immunoprecipitation assay (Fig. 4a). The binding is competed by the BMP antagonist noggin²², but not by chordin (Fig. 4a). Furthermore, using chick chordin, which is secreted more readily from the oocyte than *Xenopus* chordin, we observed that chordin is co-immunoprecipitated with Flag-Tsg in the presence of Bmp4 (Fig. 4b). These data suggest that Tsg forms a ternary complex with chordin and Bmp4, and that it may inhibit BMP function through direct binding to BMPs.

To investigate whether Tsg modifies the BMP inhibitory activity of chordin, we analysed the effect of coexpression of chordin and *Xenopus* Tsg in both ectodermal explants and whole embryos. In animal caps, a low dose of either chordin or Tsg does not block efficiently brachyury (*Bra*) induction by Bmp4; coexpression of both genes leads to a more effective inhibition of the BMP activity (Fig. 4c). In whole embryos, chordin induces a partial secondary axis when expressed at the ventral side; coexpression with *Xenopus* Tsg at a ratio of Tsg/chordin below 1/1 does not prevent the ectopic axis-induction by chordin. Instead, many embryos show a dorsolateralized phenotype with reduced trunk and tail (Fig. 4d). Notably, with an increasing Tsg/chordin ratio (2/1 or higher), the secondary axis disappears in an increasing number of embryos, and these embryos resume a typical Tsg phenotype (Fig. 2a; and data not shown). Tsg and chordin still inhibit *Bra* induction by Bmp4 in animal caps at these high ratios (not shown). The data suggest that although Tsg and chordin expression together leads to a more efficient BMP inhibition in ectodermal explants regardless of the amounts of RNA injected, the *in vivo* phenotypes depend on the relative abundance of the two genes. The exact mechanism underlying this observation remains unclear. There are at least two possibilities that are not mutually exclusive. The highly diffusible Tsg may help to diffuse the chordin protein through complex formation, which may result in dilution of the chordin function at the ventral side and inhibition of the secondary axis induction by chordin. Alternatively, other endogenous factors may interact with chordin and/or Tsg and modify their activities and the phenotypes of coexpression of chordin and Tsg.

Tsg, first identified in *Drosophila* and shown to have a function in development of the dorsal midline, was originally proposed to potentiate DPP activity^{13–15}. This hypothesis, however, has recently been challenged by the observation that a processed SOG product, which has a broader inhibitory spectrum to BMP members, rescues the *tsg* mutant phenotype, whereas DPP does not^{15,23}. This result indicates that TSG may participate in alternative processing of SOG to generate a 'supersog' to block the DPP signalling. TSG may therefore be involved in inhibiting, rather than enhancing, the DPP activity. A similar mechanism suggesting that Tsg may participate in inhibition of the BMP signalling has now been proposed to also work in vertebrates^{24,25}. Tsg stimulates chordin cleavage at a unique site, and Tsg enhances chordin activity in both *Xenopus* and zebrafish^{24,25}. Here we show that Tsg can function as a BMP antagonist in embryonic explants and interfere with BMP-dependent tissue formation in frog embryos. The activity of Tsg, however, is different from that of other BMP antagonists, as Tsg also induces defects in the anterior tissues, such as the anterior endoderm expressing the *Hex* gene and the prechordal mesoderm that expresses goosecoid. This phenotype may underlie the recent interpretation that vertebrate Tsg acts to promote BMP signalling²⁶, as it has been reported that elevated BMP expression leads to reduction of dorsal markers at the gastrula stages, which results in truncation of anterior and dorsal tissues at late stages^{4–6}. It is currently unclear, however, whether the defects induced by TSG truly reflect an enhancement of BMP signalling. Although we do not

rule out the possibility that Tsg can function both as a BMP agonist and antagonist depending on the presence or absence of other factors in specific regions of the embryos at particular developmental stages, it is also possible that the late effects are still mediated by inhibition of a BMP pathway. Several BMPs are expressed in the anterior endoderm and/or prechordal mesoderm in chick and fish^{27–29}, and in chick, BMPs may be involved in specification of prechordal mesoderm²⁷. It is possible that Tsg inhibits a different spectrum of BMPs from other BMP antagonists, either alone or by alteration of chordin specificity, thereby leading to the unique phenotype in the anterior tissues. Another possibility is that Tsg may have one or more additional functions independent of BMP signalling. Further investigation is required to understand the *in vivo* function of Tsg. □

Methods

Plasmid construction

Xenopus Tsg was cloned by low-stringency screening of a stage 28 λZapII head library (R. Harland, UC Berkeley), using the human TSG gene as the probe. The full-length *Xenopus* Tsg was subcloned by PCR strategy into the *EcoRI* and *XhoI* sites of the pCS2+ vector with the primers Tsg2-N: 5'-GCGGAATTCACCATGAAGCCCTCTTTCCTTCA-3' and the Tsg2 carboxy terminal (Tsg2-C): 5'-GCGCTCGAGTCATTAACCATACAGTT-CACGC-3'. For Flag-tagged *Xenopus* Tsg, two additional overlapping C-terminal primers were used: Flag-Tsg-C1: 5'-GCGCTCGAGTTAGTTAACTTATCATCATCATCCTTATA-3' and Flag-Tsg-C2: 5'-CTTATCATCATCATCCTTATAATCAACCATACAGTT-CACGCA-3'. Five cycles of PCR with primers Tsg2-N and Flag-Tsg-C2 were followed by a further 25 cycles after the addition of the Flag-Tsg-C1 primer. The resulting PCR product was inserted into the *EcoRI/XhoI* sites of the pCS2+ vector. For the *Xenopus* Tsg-CD2 construct, the PCR product with Tsg2-N and Flag-Tsg-C (5'-GCGCTCGAGTTAACTTATCATCATCATC-3') was digested with *PstI* and ligated with *PstI/EcoRI* fragment of the rat CD2 gene (G. Struhl, Columbia University). The ligation product was digested with *EcoRI* and inserted into the *EcoRI* site of the pCS2+ vector. The RNAs for all these constructs were synthesized on *Ascl* linearized templates, using the SP6 polymerase.

Received 10 November 2000; accepted 16 January 2001.

- Irish, V. F. & Gelbart, W. M. The decapentaplegic gene is required for dorsal-ventral patterning of the *Drosophila* embryo. *Genes Dev.* **1**, 868–879 (1987).
- Ferguson, E. L. & Anderson, K. V. Decapentaplegic acts as a morphogen to organize dorsal-ventral pattern in the *Drosophila* embryo. *Cell* **71**, 451–461 (1992).
- Wharton, K. A., Ray, R. P. & Gelbart, W. M. An activity gradient of decapentaplegic is necessary for the specification of dorsal pattern elements in the *Drosophila* embryo. *Development* **117**, 807–822 (1993).
- Dale, L., Howes, G., Price, B. M. J. & Smith, J. C. Bone morphogenetic protein 4: a ventralizing factor in early *Xenopus* development. *Development* **115**, 573–585 (1992).
- Fainsod, A., Steinbeisser, H. & De Robertis, E. M. On the function of BMP-4 in patterning the marginal zone of the *Xenopus* embryos. *EMBO J.* **13**, 5015–5025 (1994).
- Jones, C. M., Dale, L., Hogan, B. L. M., Wright, C. V. E. & Smith, J. C. Bone morphogenetic protein-4 (BMP-4) acts during gastrula stages to cause ventralization of *Xenopus* embryos. *Development* **122**, 1545–1554 (1996).
- Ferguson, E. L. & Anderson, K. V. Localized enhancement and repression of the activity of the TGF-β family member, decapentaplegic, is necessary for dorsal-ventral pattern formation in the *Drosophila* embryo. *Development* **114**, 583–597 (1992).
- Francois, V., Solloway, M., O'Neill, J. W., Emery, J. & Bier, E. Dorsal-ventral patterning of the *Drosophila* embryo depends on a putative negative growth factor encoded by the short gastrulation gene. *Genes Dev.* **8**, 2602–2616 (1994).
- Biehs, B., Francois, V. & Bier, E. The *Drosophila* short gastrulation gene prevents Dpp from autoactivating and suppressing neurogenesis in the neuroectoderm. *Genes Dev.* **10**, 2922–2934 (1996).
- Holley, S. A. et al. A conserved system for dorsal-ventral patterning in insects and vertebrates involving sog and chordin. *Nature* **376**, 249–253 (1995).
- Schmidt, J., Francois, V., Bier, E. & Kimmel, D. *Drosophila* short gastrulation induces an ectopic axis in *Xenopus*: evidence for conserved mechanism of dorsal-ventral patterning. *Development* **121**, 4319–4328 (1995).
- Piccolo, S., Sasai, Y., Lu, B. & De Robertis, E. M. Dorsolateral patterning in *Xenopus*: inhibition of ventral signals by direct binding of chordin to BMP-4. *Cell* **86**, 589–598 (1996).
- Zusman, S. B. & Wieschaus, E. F. Requirements for zygotic gene activity during gastrulation in *Drosophila melanogaster*. *Dev. Biol.* **111**, 359–371 (1985).
- Mason, E. D., Konrad, K. D., Webb, C. D. & Marsh, J. L. Dorsal midline fate in *Drosophila* embryos requires twisted gastrulation, a gene encoding a secreted protein related to human connective tissue growth factor. *Genes Dev.* **8**, 1489–1501 (1994).
- Mason, E. D., Williams, S., Grotendorst, G. R. & Marsh, J. L. Combinatorial signaling by twisted gastrulation and decapentaplegic. *Mech. Dev.* **64**, 61–75 (1997).
- Wilson, P. A. & Hemmati-Brivanlou, A. Induction of epidermis and inhibition of neural fate by Bmp-4. *Nature* **376**, 331–333 (1995).
- Sasai, Y., Lu, B., Steinbeisser, H. & De Robertis, E. M. Regulation of neural induction by the Chd and Bmp-4 antagonistic patterning signals in *Xenopus*. *Nature* **376**, 333–336 (1995).
- Hawley, S. H. B. et al. Disruption of BMP signals in embryonic *Xenopus* ectoderm leads to direct neural induction. *Genes Dev.* **9**, 2923–2935 (1995).

19. Hemmati-Brivanlou, A. & Thomsen, G. H. Ventral mesodermal patterning in *Xenopus* embryos: expression patterns and activities of BMP-2 and BMP-4. *Dev. Genet.* **17**, 78–89 (1995).
20. Schultheiss, T. M., Burch, J. B. E. & Lassar, A. B. A role for bone morphogenetic proteins in the induction of cardiac myogenesis. *Genes Dev.* **11**, 451–462 (1997).
21. Shi, Y., Katsev, S., Cai, C. & Evans, S. BMP signaling is required for heart formation in vertebrates. *Dev. Biol.* **224**, 226–237 (2000).
22. Zimmerman, L. B., De Jesus-Escobar, J. M. & Harland, R. M. The Spemann organizer signal noggin binds and inactivates bone morphogenetic protein 4. *Cell* **86**, 599–606 (1996).
23. Yu, K. *et al.* Processing of the *Drosophila* Sog protein creates a novel BMP inhibitory activity. *Development* **127**, 2143–2154 (2000).
24. Scott, I. C. *et al.* Homologues of Twisted gastrulation are extracellular cofactors in antagonism of BMP signalling. *Nature* **410**, 475–478 (2001).
25. Ross, J. J. *et al.* Twisted gastrulation is a conserved extracellular BMP antagonist. *Nature* **410**, 479–483 (2001).
26. Oelgeschlager, M., Larrain, J., Geisbert, D. & De Robertis, E. M. The evolutionarily conserved BMP-binding protein Twisted gastrulation promotes BMP signalling. *Nature* **405**, 757–763 (2000).
27. Vesque, C. *et al.* Development of chick axial mesoderm: specification of precordial mesoderm by anterior endoderm-derived TGF β family signalling. *Development* **127**, 2795–2809 (2000).
28. Dale, J. K. *et al.* Cooperation of BMP7 and SHH in the induction of forebrain ventral midline cells by prechordal mesoderm. *Cell* **90**, 257–269 (1997).
29. Hammerschmidt, M., Serbedzija, G. N. & McMahon, A. P. Genetic analysis of dorsoventral pattern formation in the zebrafish: requirement of a BMP-like ventralizing activity and its dorsal repressor. *Genes Dev.* **10**, 2452–2461 (1996).
30. Chang, C. & Hemmati-Brivanlou, A. *Xenopus* GDF6, a new antagonist of noggin and a partner of BMPs. *Development* **126**, 3347–3357 (1999).

Acknowledgements

We thank K. Lee and T. Jessell for providing the chick chordin construct; G. Struhl for providing the CD2 construct; A. Economides for technical advice on biochemical studies; and P. Wilson, D. Weinstein, E. Bell and I. Zohn for critical reading of the manuscript. We are grateful to M. O'Connor, K. Cho, L. Marsh and D. Greenspan for communicating their results before publication. This work is supported by an NIH grant to A.H.B., who is a McKnight, Merck and Klingenstein scholar.

Correspondence and requests for materials should be addressed to A.H.-B. (e-mail: brvnlou@rockvax.rockefeller.edu).

An *Arabidopsis* circadian clock component interacts with both CRY1 and phyB

Jose A. Jarillo*[†], Juan Capel*[‡], Ru-Hang Tang*[†], Hong-Quan Yang*, Jose M. Alonso*[§], Joseph R. Ecker*[§] & Anthony R. Cashmore*

* Plant Science Institute, Department of Biology, University of Pennsylvania, Philadelphia, Pennsylvania 19104-6018, USA

‡ Departamento de Biología Aplicada, Universidad de Almería, 04120 Almería, Spain

† These authors contributed equally to this work

Most organisms, from cyanobacteria to mammals, use circadian clocks to coordinate their activities with the natural 24-h light/dark cycle. The clock proteins of *Drosophila* and mammals exhibit striking homology but do not show similarity with clock proteins found so far from either cyanobacteria or *Neurospora*^{1,2}. Each of these organisms uses a transcriptionally regulated negative feedback loop in which the messenger RNA levels of the clock components cycle over a 24-h period. Proteins containing PAS domains are invariably found in at least one component of the characterized eukaryotic clocks¹. Here we describe *ADAGIO1* (*ADO1*), a gene of *Arabidopsis thaliana* that encodes a protein containing a PAS domain. We found that a loss-of-function *ado1* mutant is altered in both gene expression and cotyledon movement in circadian rhythmicity. Under constant white or blue light,

the *ado1* mutant exhibits a longer period than that of wild-type *Arabidopsis* seedlings, whereas under red light cotyledon movement and stem elongation are arrhythmic. Both yeast two-hybrid and *in vitro* binding studies show that there is a physical interaction between ADO1 and the photoreceptors CRY1 and phyB. We propose that ADO1 is an important component of the *Arabidopsis* circadian system.

In plants, the circadian clock controls daily changes of photosynthetic activities, leaf movement, cell growth and gene expression^{3,4}. Several molecular components have been described for plant circadian systems^{5–9}, including the blue light photoreceptor cryptochrome and the red light photoreceptor phytochrome, which mediate rhythm entrainment in *Arabidopsis*¹⁰. Cryptochrome also functions in both fly and mammalian circadian clocks¹¹—unexpectedly, in the latter, cryptochrome is described as having the properties of an integral component of the clock^{12–14}.

ADO1 is identical to ZTL¹⁵ and LKP1 (ref. 16), and is a member of a three-gene family together with ADO2 (LKP2)¹⁵ and ADO3 (FKF1)¹⁷ (see Supplementary Information). The PAS domain of ADO1 is similar to both that of white collar 1 (WC-1), an essential component of the *Neurospora* clock^{18,19}, and to the *Arabidopsis* proteins phototropin (NPH1)²⁰ and NPL1 (ref. 21). These domains mediate protein–protein interactions and are often associated with signalling pathways that monitor changes in light, redox potential, oxygen, small ligands and the overall energy level of the cell²². In the case of phototropin, this PAS/LOV domain binds FMN²⁰. In addition to a PAS domain, the three ADO proteins also contain an F-box domain, which may link target proteins to a ubiquitination complex (SCF complex)²³, and six kelch repeats, which are possibly implicated in protein–protein interactions²⁴ (Fig. 1; and Supplementary Information).

To define the function of ADO1, we pursued a reverse genetic approach. By using the polymerase chain reaction (PCR) to screen pools of T-DNA (transferred DNA) insertion lines, we identified an insertion in the ADO1 gene within the region encoding the kelch motifs. The T-DNA insertion adds six new codons before an in-frame stop codon occurs (Fig. 1a), which predicts a truncated protein. However, we did not detect expression of ADO1 mRNA in the mutant by northern blot analysis (Fig. 1b); therefore, *ado1* may be a null, or severely hypomorphic allele.

To examine whether ADO1, like the related WC-1 of *Neurospora*¹⁸, was affected in clock-regulated processes, we determined by northern blot analysis whether the temporal expression pattern of the endogenous *CCR2* gene was altered in the *ado1* mutant background. *CCR2* is an *Arabidopsis* gene that encodes an RNA-binding protein and exhibits circadian expression²⁵. RNA was prepared at 2-h intervals from 3-week-old wild-type and *ado1* mutant plants, which had been entrained to an 8-h photoperiod

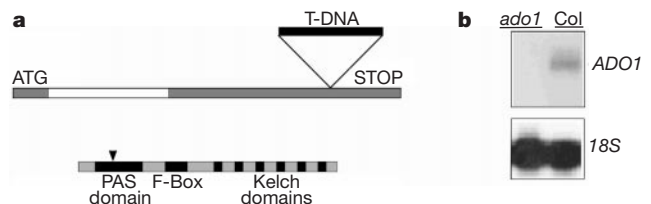


Figure 1 ADO1 gene structure and T-DNA insertion mutant. **a**, Exon structure of ADO1 gene, site of T-DNA insertion and protein domains. Filled boxes represent the two exons, open box represents the intron. The longest cDNA isolated for ADO1 was ~2,300 bp, encoding a protein of 609 amino acids. The site of insertion of the T-DNA in the *ado1* mutant is indicated, as are the conserved motifs in the ADO1 protein. Arrowhead designates C82, the codon split by the intron in the ADO1 genomic sequence. **b**, Northern blot hybridization with an ADO1 probe against RNA prepared from wild-type and *ado1* mutant seedlings. The band in the wild-type lane corresponds to a size of 2.3 kb. No hybridization signal was detected for the mutant.

§ Present address: Plant Biology Laboratory, The Salk Institute for Biological Studies, La Jolla, California 92037, USA.

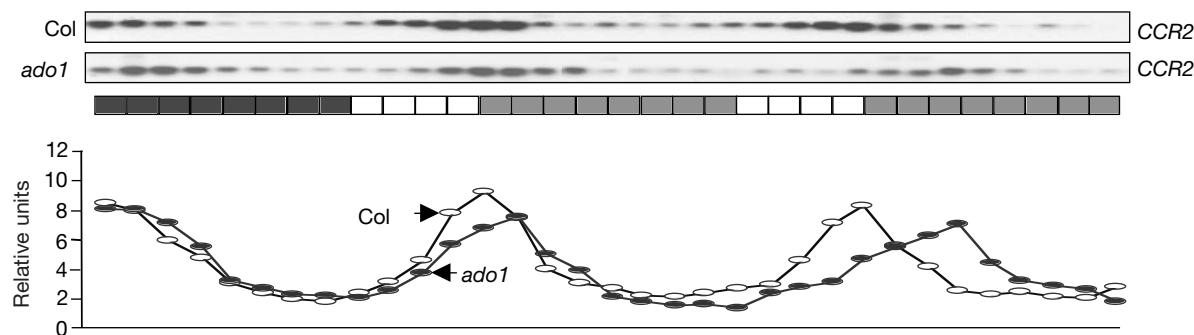


Figure 2 The periodicity of *CCR2* gene expression is lengthened in the *ado1* mutant. Top, hybridization signals for *CCR2* mRNA in the wild-type (Col) and *ado1* mutant plants. Middle, open and filled boxes indicate day and night, respectively; grey boxes represent

subjective night. Bottom, the amount of *CCR2* mRNA (relative to ribosomal RNA) is plotted for the wild-type Col and *ado1* plants.

and then removed to constant white light for a further 48 h. The period of cyclical *CCR2* gene expression was lengthened in the *ado1* mutant (Fig. 2). We observed that the period length of *CCR2* mRNA expression was roughly 24 h for the wild type, consistent with previous reports²⁵. In contrast, the period length of the *ado1* mutant was extended by about 3 h (Fig. 2). We also examined *CAB* gene expression and observed a similar change in periodicity for the *ado1* mutant (data not shown).

We examined a second, essentially unrelated, circadian process. The circadian clock controls both hypocotyl elongation and cotyledon movement in *Arabidopsis* seedlings²⁶, with circadian dysfunction disrupting these processes^{5,6,26,27}. Analysis of *ado1* mutant seedlings showed that the periodicity of cotyledon movement was lengthened by roughly 6 h in white light (Fig. 3a). Because both blue and red light also serve to entrain the circadian clock in *Arabidopsis*¹⁰, we similarly defined the periodicity of *ado1* mutant seedlings under these light conditions. We observed that the periodicity of cotyledon movements was lengthened in the *ado1* seedlings under blue light, as under white light (Fig. 3b). In marked contrast, cotyledon movement and growth of the *ado1* seedlings under red light was arrhythmic (Fig. 3c).

Because *ado1* was affected in several outputs from the circadian clock, we concluded that *ADO1* has a role, either in the clock itself or in processing light signalling to the clock. Both cryptochrome and phytochrome photoreceptors have been shown to function in entraining *Arabidopsis* circadian rhythms in response to light¹⁰, so we used the yeast two-hybrid system to define potential interactions between *ADO1* and these photoreceptors. A fusion of *ADO1* to the GAL4 DNA-binding domain (GDB) interacted with the carboxy terminus of PHYB fused to the GAL4 DNA activation domain (GAD), as shown by the production of significant levels of β -galactosidase activity (Fig. 4a, b). Similarly, a strong interaction was observed between *ADO1* and the C terminus of CRY1. No interactions were observed between any of these (*ADO1*, PHYB or CRY1) and unrelated control proteins. We extended these studies by *in vitro* binding experiments. Again, the GAD-*ADO1* fusion protein was found to bind to the C terminus of both CRY1 and PHYB (Fig. 4c). Significantly, no binding of GAD-*ADO1* was observed with the related PAS-domain-containing protein NPH1 (Fig. 4c).

Our original analysis of *ADO1* was prompted by the striking similarity of its PAS domain to that of the white collar protein WC-1, a component of the *Neurospora* circadian clock^{18,19}. Our observations, together with those indicating that mutations at the *ADO1/ZTL* locus affect the periodicity of two essentially unrelated processes—*CCR2* and *CAB* gene expression, as well as cotyledon movement¹⁵—argue strongly that *ADO1* does not simply affect output but either affects on input to the clock or is an integral component of the circadian oscillator itself. Given the similarity of

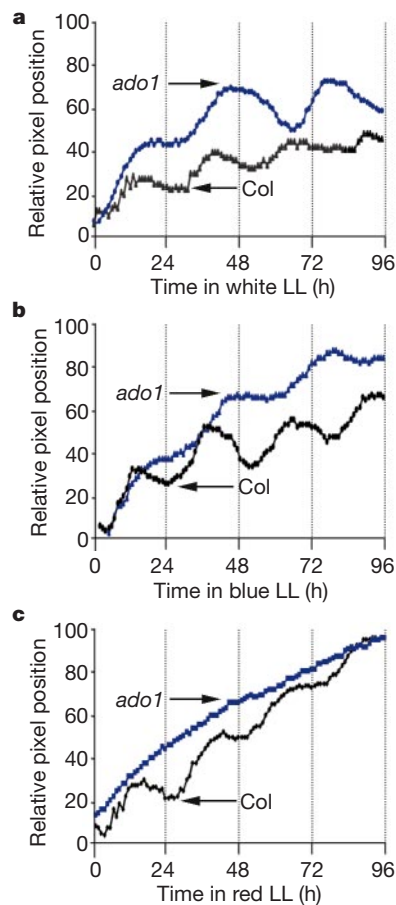


Figure 3 Cotyledon tip movement in the *ado1* mutant. Cotyledon movement was recorded for wild-type (Col) and mutant (*ado1*) seedlings grown under a 12-hour photoperiod and then transferred to constant white (a), blue (b) or red (c) light.

the *ADO1* PAS domain to the corresponding flavin-binding PAS/LOV domains of phototropin and WC-1, it might be argued that *ADO1* is a photoreceptor, as is thought to be the case for these other two PAS-domain proteins^{20,28}. An alternative viewpoint is that *ADO1* is an integral component of the *Arabidopsis* circadian clock, receiving input signals for photo-entrainment from both phytochrome and cryptochrome.

Our observation that *ADO1* physically interacts with both phytochrome and cryptochrome photoreceptors of *Arabidopsis* is consistent with both of these interpretations. Similarly, our finding that stem elongation and cotyledon movement for the *ado1* mutant

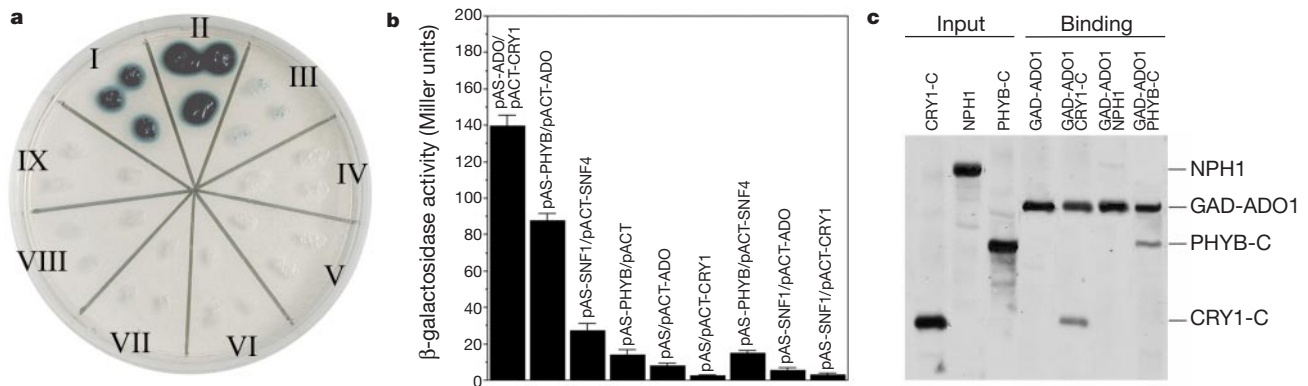


Figure 4 ADO1 interacts with both cryptochrome and phytochrome. **a**, In the yeast two-hybrid system, there is strong interaction between ADO1 and CRY1 (I), and between PHYB and ADO1 (II). A weak interaction was observed for the positive control proteins SNF1 and SNF4 (III). No interactions were observed for PHYB and pACT vector (IV), pAS vector and ADO1 (V), pAS vector and CRY1 (VI), PHYB and SNF4 (VII), SNF1 and ADO1 (VIII), or SNF1 and CRY1 (IX). **b**, Quantification of the amount of β -galactosidase produced in **a** for each

plasmid combination relative to the number of yeast cells. **c**, *In vitro* transcription/translation was used to generate polypeptides for binding studies. The autoradiograph of the SDS-PAGE gel shows the binding of *in vitro* transcription/translation products to the GAD-ADO1 fusion protein immobilized to protein A/agarose beads using antibody against GAD. Both the CRY1 C terminus and the PHYB C terminus bind to ADO1, but no binding is observed for the related PAS-domain-containing NPH1 protein.

are arrhythmic under red light does not distinguish whether ADO1 has a role in input or whether it is an integral component of the clock. The long-period rhythmicity that we observed for *ado1* under blue and white light might reflect a redundant activity of ADO2 and/or ADO3. A better understanding of the roles of these genes in the *Arabidopsis* clock should be obtained by examining mutants affected in these additional members of the ADAGIO family. Finally, irrespective of the precise function of these genes, our findings indicate that at least some components of the circadian clock in *Arabidopsis* are more similar to those in *Neurospora* than they are to those of flies and mammals. □

Methods

ADO1 sequence determination

The ADO1 genomic sequence was amplified by PCR with a 5' primer complementary to a 23-base-pair (bp) sequence in the 5' untranslated region of the ADO1 complementary DNA, 130 bp upstream of the ATG codon, and with a 3' primer complementary to a 22-bp sequence immediately downstream of the TGA codon. The amplified PCR products were sequenced using an ABI 373 sequencer. Flanking genomic sequence outside the cDNA sequence was subsequently obtained from P1 clone MSF19 (GenBank accession number AB016891).

Isolation of *ado1*, a T-DNA insertion mutant

The *ado1* mutant was identified by PCR screening of 30,000 T-DNA insertion lines (J.M.A. and J.R.E., unpublished), using oligonucleotides specific for the ADO1 gene and the T-DNA. We showed by DNA sequencing that the insertion disrupted the gene after the codon for Glu 440, adding six amino-acid codons before an in-frame stop codon occurred.

Detection of mRNA by northern hybridization

Plants were grown for 3 weeks at 22 °C under an 8-h photoperiod and then moved for 48 h to continuous white light of 120 $\mu\text{mol m}^{-2} \text{s}^{-1}$. Every 2 h, the aerial parts of at least 5 plants were pooled for RNA extractions. Total RNA (10 μg) was electrophoresed on agarose gels and transferred to Hybond N⁺ (Amersham). We analysed CCR2 gene expression with a probe prepared from expressed sequence tag 151C20T7. After northern hybridization, nylon membranes were exposed to a Bio-Rad imaging plate at room temperature. The image was visualized using a Bio-Rad FX phosphorimager. Expression levels were normalized against the signal obtained by hybridizing the same blot with an *Arabidopsis* 26S ribosomal DNA probe. The normalized values were then expressed as a proportion of the highest value obtained, and graphed. Experiments were repeated twice.

Cotyledon movement

Seeds were surface-sterilized and sown in two rows, one of mutant seed and one of wild-type, on MS medium containing 2% sucrose. The seeds were incubated in Petri dishes in a growth chamber under 12 h (80 $\mu\text{mol m}^{-2} \text{s}^{-1}$) cool white fluorescent light/12 h dark at 22 °C. After 5 d, the Petri dishes were placed before a Logitech Quickcam video camera. Cotyledon movement was recorded under continuous white, blue or red light of 10 $\mu\text{mol m}^{-2} \text{s}^{-1}$ at 22 °C. We recorded the position in pixels of one cotyledon tip per seedling for 96 h. Periods of cotyledon movements were estimated as described²⁹. The

periodicity of cotyledon movement under continuous white light (and blue light) for the mutant was about 6 h longer than that observed for wild-type. By contrast, the periodicity of CCR2 gene expression for the *ado1* mutant under constant white light was about 3 h longer than that of wild type. These differences in period extension may reflect changes in either the age of the plants and/or the entrainment conditions for the two experiments.

Yeast two-hybrid assay

We used the matchmaker yeast two-hybrid system (Clontech) for protein interaction studies. Plasmid pAS2-1 contains the yeast GBD cloning vector, whereas pACT2 contains the GAD cloning vector. To eliminate false positives, yeast cells were transformed with the bait and prey, mated, and positive colonies were selected on SD/-LEU/-TRP/-HIS/AT medium. A non-lethal β -galactosidase plate assay for yeast was achieved on agar plates by adding X-gal directly to the medium (buffered to pH 7.0), colonies were allowed to grow for 3 d, flooded with chloroform for 5 min, and then incubated at 30 °C for 24 h. Quantification was determined using a colorimetric assay with chlorophenol red β -D-galactopyranoside as the substrate according to the manufacturer's protocol. We determined the averages and standard errors for β -galactosidase activities on three colonies, and repeated this three times.

In vitro protein binding studies

DNA templates encoding the *Arabidopsis* phyB C-terminal domain (645–1172), the CRY1 C-terminal domain (490–681), the full-length NPH1 and the full-length ADO1 protein fused to GAD (GAD-ADO1) were cloned into the pBluescript SK(-) vector. These templates were transcribed with T7 polymerase and translated in the presence of [³⁵S]Met using the T'n'T-coupled *in vitro* transcription/translation reaction mix (Promega). GAD-ADO1 was attached to 20 μl of protein A/agarose beads by adding 1 μg monoclonal antibody against GAD (Santa Cruz Biotechnology). The bead-bound protein was incubated for 45 min at 4 °C in 100 μl PBS binding buffer (pH 7.4) containing complete protease inhibitor, 0.1% NP-40, 0.1% bovine serum albumin (BSA) and 1 mM phenylmethylsulphonyl fluoride (PMSF). The protein was washed twice with 1.0 ml PBS binding buffer without BSA and PMSF. For binding studies, agarose beads carrying ~3.5 fmol GAD-ADO1 were incubated with candidate binding proteins in a 60- μl reaction on a rotating wheel at 4 °C for 1 h. After centrifugation, the beads were pelleted and washed three times with the above ice-cold washing buffer. The washed pellet was resuspended in 2 $\times$ SDS sample buffer, and the proteins were analysed by SDS-PAGE and autoradiography.

Received 14 April 2000; accepted 2 January 2001.

- Dunlap, J. C. Molecular basis for circadian clocks. *Cell* **96**, 271–290 (1999).
- Lakin-Thomas, P. L. Circadian rhythms: new functions for old clock genes? *Trends Genet.* **16**, 135–142 (2000).
- Millar, A. Biological clocks in *Arabidopsis thaliana*. *New Phytol.* **141**, 175–197 (1999).
- Somers, D. The physiology and molecular bases of the plant circadian clock. *Plant Physiol.* **121**, 9–19 (1999).
- Wang, Z. & Tobin, E. Constitutive expression of the circadian clock associated 1 (CCA1) gene disrupts circadian rhythms and suppresses its own expression. *Cell* **93**, 1207–1217 (1998).
- Schaffer, R. *et al.* The late elongated hypocotyl mutation of *Arabidopsis* disrupts circadian rhythms and the photoperiodic control of flowering. *Cell* **93**, 1219–1229 (1998).
- Fowler, S. *et al.* GIGANTEA: a circadian clock-controlled gene that regulates photoperiodic flowering in *Arabidopsis* and encodes a protein with several possible membrane-spanning domains. *EMBO J.* **18**, 4679–4688 (1999).
- Sugano, S., Andronis, C., Ong, M. S., Green, R. M. & Tobin, E. M. The protein kinase CK2 is involved in regulation of circadian rhythms in *Arabidopsis*. *Proc. Natl Acad. Sci. USA* **96**, 12362–12366 (1999).

9. Park, D. H. *et al.* Control of circadian rhythms and photoperiodic flowering by the *Arabidopsis* *GIGANTEA* gene. *Science* **282**, 1579–1582 (1999).
10. Somers, D., Devlin, P. & Kay, S. Phytochromes and cryptochromes in the entrainment of the *Arabidopsis* circadian clock. *Science* **282**, 488–490 (1998).
11. Cashmore, A. R., Jarillo, J. A., Wu, Y. J. & Liu, D. Cryptochromes: blue light receptors for plants and animals. *Science* **284**, 760–765 (1999).
12. Okamura, H. *et al.* Photoc induction of *mPer1* and *mPer2* in cry-deficient mice lacking a biological clock. *Science* **286**, 2531–2534 (1999).
13. Griffin, E. A., Staknis, D. & Weitz, C. J. Light-independent role of CRY1 and CRY2 in the mammalian circadian clock. *Science* **286**, 768–771 (1999).
14. Kume, K. *et al.* mCRY1 and mCRY2 are essential components of the negative limb of the circadian clock feedback loop. *Cell* **98**, 193–205 (1999).
15. Somers, D. E., Schultz, T. F., Milnamow, M. & Kay, S. A. *ZEITLUPE* encodes a novel clock-associated PAS protein from *Arabidopsis*. *Cell* **101**, 319–329 (2000).
16. Kiyosue, T. & Wada, M. LKP1 (LOV kelch protein 1): a factor involved in the regulation of flowering time in *Arabidopsis*. *Plant J.* **23**, 807–815 (2000).
17. Nelson, D. C., Lasswell, J., Rogg, L. E., Cohen, M. A. & Bartel, B. *FKF1*, a clock-controlled gene that regulates the transition to flowering in *Arabidopsis*. *Cell* **101**, 331–340 (2000).
18. Ballario, P. *et al.* White collar-1, a central regulator of a blue light responses in *Neurospora*, is a zinc finger protein. *EMBO J.* **15**, 1650–1657 (1996).
19. Crosthwaite, S. K., Dunlap, J. C. & Loros, J. J. *Neurospora* wc-1 and wc-2: transcription, photo-responses, and the origins of circadian rhythmicity. *Science* **276**, 763–769 (1997).
20. Christie, J. M. *et al.* *Arabidopsis* NPH1: a flavoprotein with the properties of a photoreceptor for phototropism. *Science* **282**, 1698–1701 (1998).
21. Jarillo, J. A., Ahmad, M. & Cashmore, A. R. NPL1 (Accession number AF053941): a second member of the NPH serine/threonine kinase family of *Arabidopsis*. *Plant Physiol.* **117**, 719 (1998).
22. Taylor, B. L. & Zhulin, I. B. PAS domains: internal sensors of oxygen, redox potential, and light. *Microbiol. Mol. Biol. Rev.* **63**, 479–506 (1999).
23. Patton, E. E., Willems, A. R. & Tyers, M. Combinatorial control in ubiquitin-dependent proteolysis: don't skip the F-box hypothesis. *Trends Genet.* **14**, 236–243 (1998).
24. Adams, J., Kelso, R. & Cooley, L. The kelch repeat superfamily of proteins: propellers of cell function. *Trends Cell Biol.* **10**, 17–24 (2000).
25. Carpenter, C. D., Kreps, J. A. & Simon, A. E. Genes encoding glycine-rich *Arabidopsis thaliana* proteins with RNA-binding motifs are influenced by cold treatment and an endogenous circadian rhythm. *Plant Physiol.* **104**, 1015–1025 (1994).
26. Dowson-Day, M. & Millar, A. Circadian dysfunction causes aberrant hypocotyl elongation patterns in *Arabidopsis*. *Plant J.* **17**, 63–71 (1999).
27. Hicks, K. A. *et al.* Conditional circadian dysfunction of the *Arabidopsis* early-flowering 3 mutant. *Science* **274**, 790–792 (1996).
28. Ballario, P. & Macino, G. White collar proteins: PASSing the light signal in *Neurospora crassa*. *Trends Microbiol.* **5**, 458–462 (1997).
29. Ruf, T. The Lomb–Scargle periodogram in biological rhythm research: analysis of incomplete and unequally spaced time-series. *Biol. Rhythm Res.* **30**, 178–201 (1999).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank N. Bonini for comments on the manuscript. This work was supported by grants from the NIH and the DOE to A.R.C., from NSF to J.R.E., and a NATO fellowship awarded to J.C.

Correspondence and requests for materials should be addressed to A.R.C. (e-mail: cashmore@sas.upenn.edu).

A metabolic enzyme for S-nitrosothiol conserved from bacteria to humans

Limin Liu[†], Alfred Hausladen[†], Ming Zeng[†], Loretta Que[†], Joseph Heitman[‡] & Jonathan S. Stamler^{*†§}

^{*} Howard Hughes Medical Institute, [†] Department of Medicine, Pulmonary and Cardiology Divisions, and [§] Department of Biochemistry, Duke University Medical Center, Durham, North Carolina 27710, USA
[‡] Department of Genetics, Duke University Medical Center, Durham, North Carolina 27710, USA

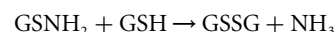
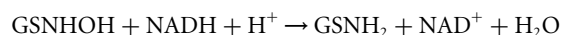
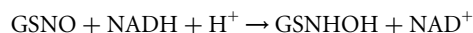
Considerable evidence indicates that NO biology involves a family of NO-related molecules and that S-nitrosothiols (SNOs) are central to signal transduction and host defence^{1–5}. It is unknown, however, how cells switch off the signals or protect themselves from the SNOs produced for defence purposes. Here we have purified a single activity from *Escherichia coli*, *Saccharomyces cerevisiae* and mouse macrophages that metabolizes S-nitrosoglu-

tathione (GSNO), and show that it is the glutathione-dependent formaldehyde dehydrogenase. Although the enzyme is highly specific for GSNO, it controls intracellular levels of both GSNO and S-nitrosylated proteins. Such 'GSNO reductase' activity is widely distributed in mammals. Deleting the reductase gene in yeast and mice abolishes the GSNO-consuming activity, and increases the cellular quantity of both GSNO and protein SNO. Furthermore, mutant yeast cells show increased susceptibility to a nitrosative challenge, whereas their resistance to oxidative stress is unimpaired. We conclude that GSNO reductase is evolutionarily conserved from bacteria to humans, is critical for SNO homeostasis, and protects against nitrosative stress.

The current perspective on the breakdown of SNOs *in vivo* is shaped by the facile decomposition of these compounds *in vitro* by ascorbate, thiols and copper ions^{6–8}. Although a number of enzymes can decompose SNOs *in vitro*, including the thioredoxin system⁹, glutathione peroxidase¹⁰, γ -glutamyl transpeptidase¹¹, xanthine oxidase¹² and glutathione-dependent formaldehyde dehydrogenase (GS-FDH)¹³, none has been shown to regulate levels of any endogenous SNO or to be involved in any NO or SNO-mediated response. The activity of GS-FDH has, however, been shown to be particularly robust¹³.

We previously observed that *E. coli* rapidly degrades SNOs, and concluded that intracellular GSNO was being metabolized³. We attributed part of the GSNO-metabolizing activity to 'SNO lyases', which cleave the SNOs homolytically to NO (ref. 14). Only a fraction of the added SNO, however, could be recovered as nitrite/nitrate, suggesting that there are other reductive routes for SNO metabolism^{3,14}. In support of this, we detected a GSNO-consuming activity in *E. coli* lysates that depended on NADH (nicotinamide adenine dinucleotide) but not O₂ (Fig. 1a). We purified this activity to homogeneity and identified it as glutathione-dependent formaldehyde dehydrogenase (GS-FDH) by amino-terminal sequencing (see Methods).

The catalytic rate constant/Michaelis constant (k_{cat}/K_M) ratios for GSNO were near diffusion-limited rates (Fig. 1b). The enzyme was highly specific for GSNO: no activity was seen for S-nitrosocysteine and S-nitrosohomocysteine, whereas the SNO derivatives of the glutathione (GSH) metabolites cysteinyl-glycine and γ -glutamyl-cysteine were degraded at ~1% of the rate of GSNO. Ammonia (NH₃) and glutathione disulphide (GSSG) were main products of the *E. coli* enzyme, and the yields depended on GSH, consistent with the following scheme (see also Methods, 'Products of *E. coli* GSNO reductase'):



We previously developed a mouse macrophage (RAW 264.7) model of nitrosative stress, characterized by the accumulation of SNOs to toxic levels after activation by interferon- γ (IFN- γ) and lipopolysaccharide (LPS), and ultimately apoptotic cell death¹⁵. As a first step to understand the fate of SNOs, cytosolic constituents were separated by size. 'High-mass' S-nitrosothiols, with a larger relative molecular mass (M_r) than 5,000 (5K), were present in substantial amounts in the cell lysate (Fig. 2a), whereas 'low-mass' SNOs (< 5K) were not detectable (Fig. 2a) (limit of sensitivity 1 pmol GSNO⁴). As GSNO (~300 pmol per 10⁶ cells) formed readily in the extracellular medium of IFN- γ /LPS-treated RAW 264.7 cells incubated with GSH¹⁶, and GSH is the predominant source of intracellular thiol, we reasoned that GSNO was either rapidly exported or metabolized. Rates of SNO accumulation in the medium of RAW 264.7 cells were very slow and at the limits of detection (ref. 16; and data not shown), whereas GSNO was metabolized quickly when it was incubated with extracts from either resting or cytokine-activated

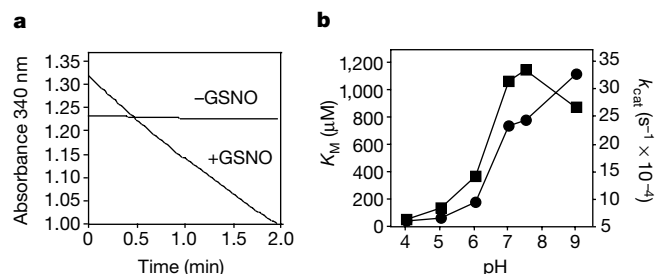


Figure 1 *Escherichia coli* GSNO reductase. **a**, Consumption of GSNO by *E. coli* lysates. A crude extract (500 $\mu g\ ml^{-1}$; strain RK4936) was incubated anaerobically with 0.2 mM NADH in the absence (–) or presence (+) of 0.15 mM GSNO, and NADH consumption (absorbance at 340 nm) was followed over time. The experiment was performed anaerobically to eliminate nonspecific NADH oxidation by diaphorases present in *E. coli* lysates. **b**, K_M (circles) and k_{cat} (squares) values of purified *E. coli* GSNO reductase (under aerobic conditions). (K_M is 30 μM at pH 4). Data were fitted to the Michaelis–Menten equation using Sigmaplot.

macrophages (data not shown). As in *E. coli*, the GSNO-metabolizing activity required NADH and was ineffective at degrading alternative SNOs.

GSNO-metabolizing activity of RAW 264.7 cell lysates was recovered in a single fraction from Mono Q anion exchange chromatography (Fig. 2b). Purification over a 5'AMP-Sepharose column yielded a single band of about 43K in a silver-stained SDS polyacrylamide gel electrophoresis (PAGE) gel (Fig. 2c). Trypsin digestion followed by peptide-mass analysis identified the protein as mouse GS-FDH or alcohol dehydrogenase class III (ADH III) (Table 1), the homologue of the bacterial GSNO reductase. Kinetic analysis (in the absence of added glutathione) (Fig. 2d) gave a K_M of 20 μM , an estimated K_{cat} of 5,600 min^{-1} and a stoichiometry of GSNO to NADH of 1:1 (data not shown). Macrophage GSNO reductase also oxidized the formaldehyde-GSH adduct, S-hydroxymethylglutathione, previously thought to be the major enzyme substrate, but the specific activity was only about 6% of the GSNO-reducing activity (data not shown). Thus, GSNO is the preferred substrate of this enzyme. In sum, the GSNO reductase activity of mouse macrophages is high and specific toward GSNO, and seems to constitute a metabolic pathway (Fig. 2a) that may protect the cells against nitrosative stress.

We detected NADH-dependent GSNO reductase activity in mouse SVEC4-10 endothelial cells, mouse SV40LT-SMC smooth muscle cells and mouse liver cells (data not shown); and in all instances it accounted for most of the GSNO-consuming activity. Moreover, GSNO reductase activity was widespread in various human cell lines (Table 2).

To establish that GSNO reductase activity is a primary means for degrading GSNO in tissues, we monitored GSNO decomposition in mouse liver extracts supplemented with various redox cofactors (ascorbate, glutathione) that have been proposed to decompose GSNO^{7,8}, and compared these activities in wild-type mice with those from littermates in which we had knocked-out the GSNO reductase (GS-FDH^{-/-}). These studies show that the NADH-dependent GSNO reductase activity dominates alternative decomposition reactions that operate in simplified *in vitro* systems (Fig. 2e). Comparison of hepatocytes from GS-FDH^{-/-} mice and wild-type littermates, at 14 and 40 h after *ex-vivo* treatment with LPS and cytokines, showed that levels of total SNO, but not of alternative NO complexes¹⁵, were about 50% higher in cells deficient in GSNO reductase and that most of this SNO was bound to protein. Moreover, whereas GSNO was barely detectable in wild-type hepatocytes, the levels were increased by 60–175% in the GS-FDH^{-/-} cells (data not shown).

We previously showed that yeast is an excellent model for studying biochemical correlates of nitrosative stress and provides an

Table 1 Peptide-mass matches for the mouse macrophage GSNO reductase and GS-FDH

Measured	Calculated	Difference (%)	Residues
1700.538	1700.735	0.0116	234–248
1899.711	1899.867	0.0082	232–248
1919.757	1919.912	0.0081	85–101
2010.822	2010.907	0.0082	169–188
2304.175	2304.237	0.0027	10–31
2622.330	2622.268	–0.0024	136–159
2857.594	2857.509	–0.0030	285–312
3154.661	3154.567	–0.0062	342–369

A search of the EMBL data base (GPMW program) with mass data obtained for 31 peptides found mouse GS-FDH (P28474) to be the only high-score hit. Shown above are 8 peptide-mass matches, which cover 42% of the predicted protein sequence, as indicated in bold:

1 mangvirckA AVAWAGKPL SIEIEVAPP Kahevrikil atavchtday tlsqrpepgc
61 fpvilghega givesvgegv tkllAGDTVI PLYIPQCCEC Kfcilnptnl cqkirtvtgk
121 glmpdgtstrf tckgkSVFHF MGTSTFSEYI VVADISVAKI dapsapldkVC LLGCGISTGY
181 GAAVNTAKve pgstcavfgl ggvglavimg ckvagasrii gidinkdkfa KAKEFGASEC
241 ISPQDFSKsi gevlvemtgd gvdysfecig nvkvmsale aahkGWGVSV VVGVAASGEE
301 ISTRPFLVLT GRtwkgtafg gwksvesvpk lvseymskik kVDEFVTGNL SFDQINQAFD
361 LMHSGDSIRT vIkM.

opportunity to dissect genetically the function of NO-related genes¹⁷. After establishing that mouse GS-FDH shares over 60% sequence identity with the yeast protein, and verifying that haploid cells of yeast strain Y190 cells have robust GSNO reductase activity (Fig. 3b), we replaced the entire open reading frame of the yeast GS-FDH gene¹⁸ (*SFA1/YDL168w*) with the dominant selectable cassette KanMX2 (resistant to G418) or hphMX (resistant to hygromycin) (Fig. 3a). Comprehensive analyses of *sfa1* (*gs-fdh*) mutant yeast (four G418^r and four hygromycin^r clones) showed that they had lost all GSNO reductase activity, but had normal levels of intracellular glutathione (Fig. 3b; data not shown). Furthermore, tetrad analysis in the *sfa1* mutant diploid strain JK93d showed co-segregation of antibiotic resistance and loss of GSNO reductase activity at 2:2 ratio in all four complete tetrads studied.

When wild-type Y190 yeast cells were treated with GSNO, the yeast accumulated only low levels of intracellular, high-mass (retained by a 5K cut-off filter) SNO (Fig. 3c). In contrast, the SNO levels were more than 11-fold higher in isogenic *sfa1* mutants (both *sfa1* Δ : G418^r and *sfa1* Δ : hygromycin^r; Fig. 3c) and substantial amounts of low-mass SNO were detected (Fig. 3c). This low-mass SNO pool was metabolized completely by adding purified GSNO reductase (data not shown), thus identifying it as GSNO. Levels of SNO were also higher in *sfa1* mutants than in wild-type cells after incubation with the NO donor DETA NONOate (data not shown). Transformation with plasmids bearing the wild-type *SFA1* gene constructs into the mutant cells partially normalized SNO levels (both protein and low-mass SNO) after an NO challenge (data not shown).

In parallel studies of GSNO reductase function, growth of *sfa1* mutant cells (both G418^r and hygromycin^r clones) was inhibited by GSNO at concentrations that had little effect on wild-type Y190 cells (Fig. 3d). In contrast, these *gs-fdh* mutant cells were no more sensitive than wild-type cells to the toxic effects of H₂O₂ (Fig. 3e). Together, these data show that GSNO reductase is essential for protection against nitrosative stress in yeast but offers no resistance to oxidative stress, and that protection is conferred by metabolizing GSNO, thereby shielding vital proteins from what are presumably toxic levels of S-nitrosylation.

GS-FDH/ADH III has been identified in many bacteria, yeasts, plants and animals^{18–21}. It is obligate for growth in methylotrophic

Table 2 GSNO metabolism (%) by human cell lines*

	No lysate	HeLa	A549	THP-1
Buffer only	6.2	29.9	14.7	18.8
+ NADH	0.0	84.2	72.3	90.8
+ NADPH + GSH + ASA	0.7	35.4	43.9	22.6

* GSNO (200 μM) was incubated with 0 or 4 $\mu g\ \mu l^{-1}$ of the cell lysates as described in Methods. GSNO metabolizing activity is shown as the percentage of the starting material that is consumed. ASA, ascorbate; GSH, glutathione.

yeast (*Pichia pastoris*) and bacteria (*Paracoccus denitrificans*) and in photosynthetic bacteria (*Rhodospirillum rubrum*)^{22–24} when methanol is the only carbon source; methanol is then metabolized to the antiseptic agent formaldehyde. In methylotrophic microorganisms, GS-FDH is induced by methanol to prevent formaldehyde accumulation^{22,24}. The physiological significance of formaldehyde oxidation by GS-FDH is less clear in other microorganisms and animals²¹.

ADH was first implicated in metabolism of nitrosocompounds more than 20 years ago (named 'nitrosoreductase')²⁵; however, its activity has been viewed as a nonspecific function of hepatocytes. A GSNO-metabolizing activity of hepatic ADH III (GS-FDH) was originally identified¹³ through studies on the detoxification of a mutagenic *N*-nitrosocompound (1,3-dimethyl-2-cyano-1-nitroso-guanidine). Here we emphasize that GS-FDH activity towards

GSNO is much greater than towards the formaldehyde-GSH adduct *S*-hydroxymethylglutathione (Figs 1b and 2d; see also refs 13, 19), and that it is found in many microbial and mammalian cells (Figs 1–3, and Table 2). Moreover, substrate GSNO is relatively ubiquitous: it is a main product of NOS (ref. 5), it constitutes a large reservoir of NO-bioactivity in extracellular fluids (which do not contain GSNO reductase)^{2,26}, and it is formed in microorganisms subjected to nitrosative challenge (Fig. 3c). Endogenous SNOs have also been isolated from nematodes and insects^{27,28}. These observations may help to explain why GS-FDH is so widely distributed and highly conserved in nature.

We previously discovered NO reductase/NO oxygenase reactions

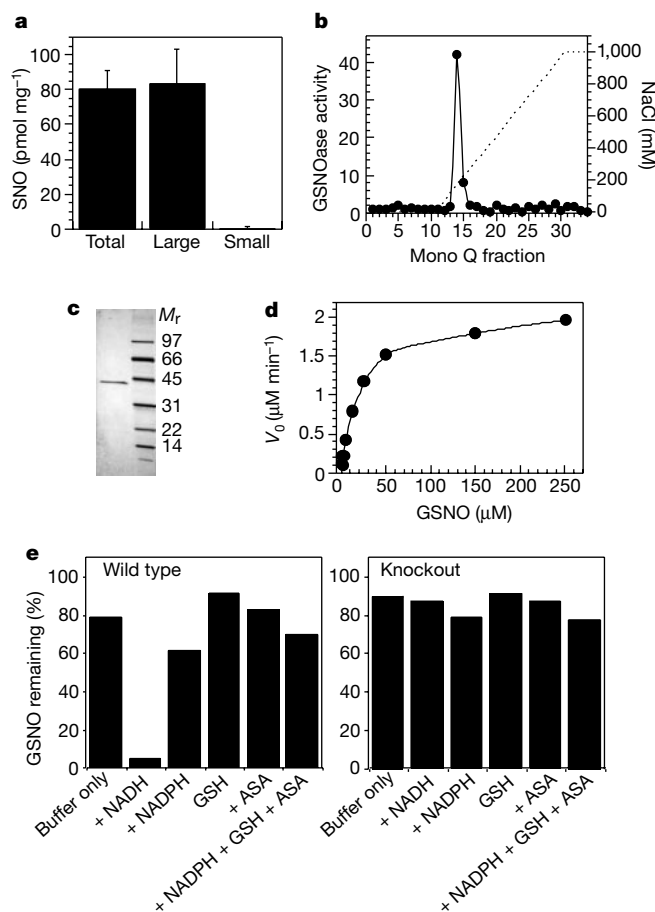


Figure 2 Mouse GSNO reductase. **a**, *S*-nitrosothiols in RAW 264.7 cells after cytokine-inducible NO synthase activation. Cells were stimulated with IFN- γ and LPS for 18 h as described³⁰. SNO levels (normalized to protein content of the lysate) in the whole lysate (Total), in the fraction that passed through a Bio-Gel P-6 column (Large), and in the fraction passing through a 5K cut-off ultrafiltration membrane (Small). Data are the mean $\pm$ s.e. of three experiments. **b**, Purification of the GSNO reductase from RAW 264.7 cells. Anion exchange chromatography (Mono Q). Solid line and circle, GSNO reductase activity; dashed line, NaCl concentration. **c**, SDS-PAGE and silver stain of the purified GSNO reductase. Lane 1, 15 ng of the GSNO reductase after purification with a 5' AMP affinity column; lane 2, molecular weight standards. **d**, Kinetic analysis of the purified macrophage GSNO reductase. K_M , $20 \pm 1 \mu\text{M}$; V_{max} , $2.09 \pm 0.03 \mu\text{M min}^{-1}$; k_{cat} , $5,600 \pm 100 \text{ min}^{-1}$ (Regression wizard, Sigma plot software). **e**, GS-FDH is a major GSNO-metabolizing activity. GSNO (200 μM) was incubated with 0 or 3 $\mu\text{g } \mu\text{l}^{-1}$ liver homogenate from wild-type and GS-FDH-deficient (knockout) mice, supplemented with NADH (300 μM), NADPH (300 μM), GSH (2 mM), ascorbic acid (ASA, 500 μM), or combinations thereof. GSNO remaining at 5 min is shown. Data are the mean of two independent experiments.

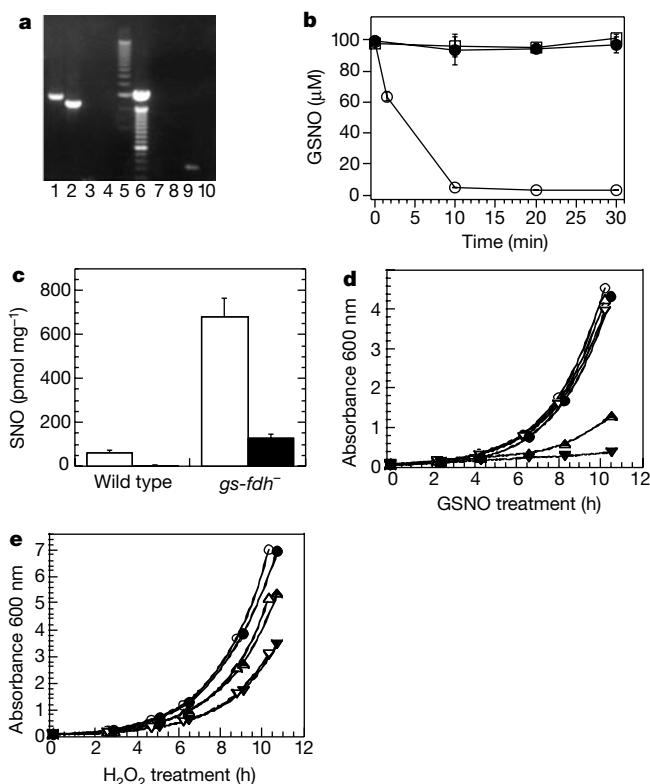


Figure 3 GSNO reductase in *S. cerevisiae* protects from nitrosative stress. **a**, Deletion of the GSNO reductase gene encoding *SFA1* (GS-FDH) in haploid Y190 cells. The entire open reading frame (ORF) of *GS-FDH* was replaced by either hphMX (lanes 1 and 7) or KanMX2 (lanes 2 and 8) cassettes. Primer pairs FDH754se/Kanas (lanes 1–4) and FDH754se/FDH1099as (lanes 7–10) were used in PCR to amplify targeted and wild-type *GS-FDH* fragments, respectively. Lanes 3 and 9, wild-type Y190 cells; lanes 4 and 10, no DNA template control; lanes 5 and 6, 500-bp and 100-bp DNA size ladders (Gibco), respectively. **b**, *GS-FDH* is required for GSNO metabolism. GSNO was incubated in buffer (square) or lysates (0.57 mg ml⁻¹) of wild-type (open circle) or mutant *gs-fdh* (G418^r, filled circle) yeast cells for indicated times. Data (mean $\pm$ s.e.) are the mean of two independent experiments. **c**, Increased levels of intracellular *S*-nitrosothiols in *gs-fdh* mutant cells after GSNO treatment. Mid-log phase (absorbance 600 nm = 0.4–0.6) cells were cultured in the presence of 5 mM GSNO at 30 °C for 2 h. SNO signals in the whole lysate (open) and the fraction that passed through a 5K cut-off membrane (filled) were normalized against whole-cell lysate protein content. Data are the mean $\pm$ s.e. from three independent experiments. **d**, Hypersensitivity of *gs-fdh* mutants to GSNO. Mid-log phase Y190 wild-type (open) or *gs-fdh* mutant (filled) cells (absorbance 600 nm = 0.05) were diluted (to absorbance 600 nm = 0.05) and cultured in YPD medium supplemented with 0 mM (circle), 2 mM (upright triangle) and 5 mM (inverted triangle) GSNO. Growth was measured as absorbance at 600 nm. Data shown are representative of three similar experiments. **e**, Lack of hypersensitivity of *gs-fdh* mutants to H₂O₂. Cells (Y190, open; *gs-fdh*, filled) were cultured in the presence of 0 mM (circle), 0.2 mM (upright triangle) and 0.6 mM (inverted triangle) H₂O₂. Growth was measured as absorbance at 600 nm. Data shown are representative of two similar experiments.

of microbial flavohaemoglobins^{14,17} (yielding N_2O and NO_3^-), which protect *E. coli* and yeast against NO, and which are reminiscent of the superoxide detoxification reactions of superoxide dismutase (yielding O_2 and H_2O_2). We have also identified an NO-metabolizing activity of nematode haemoglobin²⁷ that bears resemblance to the NO-consuming activity reported for mammalian peroxidases²⁹. Our original studies in bacteria also revealed a strong constitutive GSNO reductase activity^{3,14} that we now show is conserved throughout phylogeny and identify as GS-FDH. This enzymatic function might be considered analogous to that of NADH peroxidase. Although the SNO consumption activity of GS-FDH is highly specific for GSNO, it may have a broader function in alleviating stress imposed by nitrosants (for example, N_2O_3 , NO/O_2^- -related species), which will react with glutathione to form GSNO, and by protein SNOs, which are in equilibrium with GSNO.

Our findings have revealed in microorganisms the emergence—evidently well before O_2 accumulated in the atmosphere—of specific enzymatic systems that protect against NO and GSNO, and the conservation of such pathways from bacteria to animals. These data provide strong support for the concept of 'nitrosative stress', the molecular correlate of which is the NO-related covalent modification of proteins^{3,15,17}, and emphasize that many classes of enzymes may regulate NO bioactivity marked for signalling or defence purposes. The idea that the nitric oxide system is subject to regulation at several levels may provide a new perspective on NO biology in health and disease. □

Methods

Characterization of GSNO reductase from *E. coli*

The GSNO-metabolizing activity from *E. coli* was purified from 8 l of stationary-phase cells. A 100,000g supernatant in 20 mM BisTrisPropane, pH 7 was applied to a 5×40 cm Q-Sepharose column and eluted with a linear NaCl gradient in 20 mM BisTrisPropane, pH 7. Active fractions were pooled, adjusted to 1 M $(NH_4)_2SO_4$, and applied to a 2.5×20 cm column of butyl-Sepharose. Elution was by a decreasing $(NH_4)_2SO_4$ gradient from 1 to 0 M. After gel filtration on a HiPrep 16/10 desalting column (Pharmacia), the enzyme was purified on a MonoQ column. Active fractions were applied to a 1.6×10 cm column of AMP-Sepharose, washed with 0.15 M NaCl, and eluted with 20 mM NAD^+ . The protein was judged to be pure by SDS-PAGE (yield 0.52 mg). Limited N-terminal sequencing after blotting onto a PVDF membrane yielded a sequence of 30 amino acids (MKSRAAFAFAPGKPLEIVEIDVAPPKKGEV), which identified the protein as the GSH-dependent formaldehyde reductase.

Kinetic parameters of the *E. coli* enzyme

Assays shown in Fig. 1 were performed using 12 nM enzyme, 0.2 mM NADH and 10–500 μ M GSNO. The buffers used, each containing 0.1 M diethylenetriaminepentaacetic acid were 100 mM sodium acetate (pH 4–5), 100 mM MES (pH 6), 100 mM BisTrisPropane (pH 7) and 100 mM Tris (pH 7.5–9).

Products of *E. coli* GSNO reductase

Glutathione, ammonia, hydroxylamine, nitrous oxide, nitrite and nitrate were determined by standard spectrophotometric and mass spectroscopic methods^{3,8,13,14} in reactions containing purified GS-FDH (10 nM) at pH 4–9, and with GSH concentrations ranging from 0 to twofold that of GSNO. Nitrite, NO_3^- , NH_2OH and N_2O were not detected. Glutathione disulphide yields ranged from 30 to 100%, and NH_3 ranged from 50 to 100% that of the added GSNO, depending on GSH concentration and pH. NH_3 yields were generally higher at low pH in the presence of glutathione. At a ratio of GSNO:GSH equal to 1, pH 5, GSNO was converted stoichiometrically to GSSG and NH_3 and the stoichiometry of NADH to GSNO was about 2:1 (see proposed scheme in text).

Characterization of GSNO reductase from RAW 264.7 cells

The GSNO reductase activity in the cell lysate from mouse RAW 264.7 cells was precipitated by 45–75% ammonium sulphate. After dialysis against 20 mM Tris-HCl (pH 8.0), the proteins were fractionated on a Mono Q column with increasing concentrations of NaCl. The GSNO reductase activity was purified with a 5' AMP-Sepharose 4B affinity column (Amersham Pharmacia) in 20 mM phosphate buffer (pH 7.2). After SDS-PAGE and Coomassie blue staining, the purified protein was digested by trypsin, and the resulting peptides were analysed by mass spectrometry (Duke Univ. Protein Sequencing Facility). Initial velocities of the purified protein (Fig. 2d) were determined by GSNO-dependent NADH consumption using 30 ng ml^{-1} of the enzyme, 150 μ M NADH and varying concentrations of GSNO.

GSNO reductase activities of mammalian cells

Mouse macrophage RAW 264.7, endothelial SVEC4-10, smooth muscle SV40LT-SMC cell

lines, and human HeLa, epithelial A549 and monocyte THP-1 cell lines were obtained from the ATCC. Mouse cells were homogenized by sonication in a solution containing 20 mM Tris-HCl (pH 8.0), 0.5 mM EDTA, 0.1% NP-40 and 1 mM phenylmethylsulphonyl fluoride (PMSF). To detect GSNO-metabolizing activity, 0.85–1 mg ml^{-1} lysate was incubated with 100 μ M GSNO in reaction buffer (20 mM Tris-HCl, pH 8.0, 0.5 mM EDTA) with 0 or 200 μ M NADH at room temperature for various times. Conditions for human cell lines are as described for mouse hepatocytes (see below). SNO levels in the reaction mixture were measured by the Saville assay³. The GSNO reductase activity was also measured by GSNO-dependent NADH consumption using either absorbance (340 nm) or fluorescence (340/455 nm).

GSNO reductase activity in hepatocytes

Generation of GS-FDH^{-/-} mice will be described elsewhere. Liver tissues from wild-type and GS-FDH-deficient mice were homogenized in an enzyme reaction buffer (phosphate-buffered saline, GIBCO 14190-144, supplemented with 0.1% NP-40 and 1 mM PMSF). GSNO (200 μ M) was incubated with 0 or 3 μ g μ l⁻¹ of the liver homogenate in buffer supplemented with NADH (300 μ M), NADPH (300 μ M), GSH (2 mM), ascorbic acid (ASA, 500 μ M), or combinations thereof at 37 °C for 5 min. After precipitating the reaction mixtures with trichloroacetic acid (8.3% final concentration), we diluted the supernatants threefold in the buffer and measured SNO levels by the Saville assay³.

NO synthase activation and SNOs in mouse hepatocytes

Hepatocytes of wild-type mice and their GS-FDH^{-/-} littermates were collected after perfusion of the livers with Liberase (Roche), and purified by repeated differential sedimentation at 50g, followed by centrifugation over a 30% Percoll solution. We plated hepatocytes (over 95% viability by trypan blue exclusion) at a density of 2×10^4 cells per cm^2 in gelatin-coated flasks, and incubated them for 24 h. The cells were then cultured in the presence of recombinant mouse tumour-necrosis factor- α (500 units ml^{-1}), interferon- γ (100 units ml^{-1}), interleukin-1 β (200 units ml^{-1}) and LPS (10 μ g ml^{-1} ; *E. coli* O128:B12; Sigma) for 14 and 40 h. Nitrosylation levels in hepatocytes were measured as described¹⁵.

Deletion of the GSNO reductase gene from *S. cerevisiae*

The entire open reading frame of the GSNO reductase gene (*SFA1*) (GenBank accession number Z74216) in haploid Y190 strain (Clontech) was deleted using either KanMX2 or hphMX cassettes¹⁷. We used primers FDHKOse2 (5'-GTCCGCCCTACTGTGGTAAACCTATTAAAGTCATTGCTGATATCAAGCTTGCCTCGTC-3') and FDHKOas2 (5'-CCTATTTTATTCATCAGACTTCAAGACGGTTCTTAAGCAGTCGACACTGGATGGCGGCG-3') to amplify and add GS-FDH sequences to both ends of the cassettes by polymerase chain reaction (PCR). Cells stably transformed with KanMX2 and hphMX were selected by their resistance to G418 (200 mg ml^{-1}) and hygromycin (200 mg ml^{-1}), respectively. Replacement of the GS-FDH gene by KanMX2 or hphMX in the yeast genome was confirmed by PCR detection of GS-FDH-KanMX2 or GS-FDH-hphMX fusion fragments with primers FDH754se (5'-AAGGGAAGGACATCCGTGTG-3') and Kanas (5'-GTCGACACTGGATGGCGGCG-3'). We used primers FDH754se and FDH1099as (5'-TACGATACCGGCTCCTTCGT-3') to amplify the wild-type GS-FDH fragment. Replacement of the *SFA1* gene with Kan MX2 and hphMX was also carried out in the diploid yeast strain JK93d. Once the cassette had targeted one of the two alleles, diploid cells resistant to either G418 or hygromycin were induced to sporulate. Haploid clones of wild-type and mutant cells obtained by tetrad dissection revealed that loss of GSNO reductase activity co-segregated with resistance to antibiotics (G418^r or hyg^r); a 2:2 ratio was seen in all four complete tetrads studied.

Received 7 December 2000; accepted 5 January 2001.

- De Groote, M. A., Testerman, T., Xu, Y., Stauffer, G. & Fang, F. C. Homocysteine antagonism of nitric oxide-related cytotoxicity in *Salmonella typhimurium*. *Science* **272**, 414–417 (1996).
- Gaston, B. et al. Endogenous nitrogen oxides and bronchodilator S-nitrosothiols in human airways. *Proc. Natl Acad. Sci. USA* **90**, 10957–10961 (1993).
- Hausladen, A., Privalle, C. T., Keng, T., DeAngelo, J. & Stamler, J. S. Nitrosative stress: activation of the transcription factor OxyR. *Cell* **86**, 719–729 (1996).
- Mannick, J. B. et al. Fas-induced caspase denitrosylation. *Science* **284**, 651–654 (1999).
- Mayer, B. et al. A new pathway of nitric oxide/cyclic GMP signaling involving S-nitrosoglutathione. *J. Biol. Chem.* **273**, 3264–3270 (1998).
- Askew, S. C., Barnett, D. J., McAninly, J. & Williams, D. L. H. Catalysis by Cu^{2+} of nitric oxide release from S-nitrosothiols. *J. Chem. Soc. Perkin Trans. 2*, 741–745 (1995).
- Scorza, G., Pietraforte, D. & Minetti, M. Role of ascorbate and protein thiols in the release of nitric oxide from S-nitroso-albumin and S-nitroso-glutathione in human plasma. *Free Radicals Biol. Med.* **22**, 633–642 (1997).
- Singh, S. P., Wishnok, J. S., Keshive, M., Deen, W. M. & Tannenbaum, S. R. The chemistry of the S-nitrosoglutathione/glutathione system. *Proc. Natl Acad. Sci. USA* **93**, 14428–14433 (1996).
- Nikitovic, D. & Holmgren, A. S-nitrosoglutathione is cleaved by the thioredoxin system with liberation of glutathione and redox regulating nitric oxide. *J. Biol. Chem.* **271**, 19180–19185 (1996).
- Hou, Y., Guo, Z., Li, J. & Wang, P. G. Seleno compounds and glutathione peroxidase catalyzed decomposition of S-nitrosothiols. *Biochem. Biophys. Res. Commun.* **228**, 88–93 (1996).
- Hogg, N., Singh, R. J., Konorev, E., Joseph, J. & Kalyanaraman, B. S-Nitrosoglutathione as a substrate for gamma-glutamyl transpeptidase. *Biochem. J.* **323**, 477–481 (1997).
- Trujillo, M., Alvarez, M. N., Peluffo, G., Freeman, B. A. & Radi, R. Xanthine oxidase-mediated decomposition of S-nitrosothiols. *J. Biol. Chem.* **273**, 7828–7834 (1998).
- Jensen, D. E., Belka, G. K. & Du Bois, G. C. S-Nitrosoglutathione is a substrate for rat alcohol dehydrogenase class III isoenzyme. *Biochem. J.* **331**, 659–668 (1998).
- Hausladen, A., Gow, A. J. & Stamler, J. S. Nitrosative stress: metabolic pathway involving the flavohemoglobin. *Proc. Natl Acad. Sci. USA* **95**, 14100–14105 (1998).

15. Eu, J. P., Liu, L., Zeng, M. & Stamler, J. S. An apoptotic model for nitrosative stress. *Biochemistry* **39**, 1040–1047 (2000).
16. Akaike, T. *et al.* Nanomolar quantification and identification of various nitrosothiols by high performance liquid chromatography coupled with flow reactors of metals and Griess reagent. *J. Biochem. (Tokyo)* **122**, 459–466 (1997).
17. Liu, L., Zeng, M., Hausladen, A., Heitman, J. & Stamler, J. S. Protection from nitrosative stress by yeast flavohemoglobin. *Proc. Natl Acad. Sci. USA* **97**, 4672–4676 (2000).
18. Wehner, E. P., Rao, E. & Brendel, M. Molecular structure and genetic regulation of SFA, a gene responsible for resistance to formaldehyde in *Saccharomyces cerevisiae*, and characterization of its protein product. *Mol. Gen. Genet.* **237**, 351–358 (1993).
19. Guthrie, W. G., Holmquist, B. & Vallee, B. L. Purification, characterization, and partial sequence of the glutathione-dependent formaldehyde dehydrogenase from *Escherichia coli*: a class III alcohol dehydrogenase. *Biochemistry* **31**, 475–481 (1992).
20. Shafqat, J. *et al.* Pea formaldehyde-active class III alcohol dehydrogenase: common derivation of the plant and animal forms but not of the corresponding ethanol-active forms (classes I and P). *Proc. Natl Acad. Sci. USA* **93**, 5595–5599 (1996).
21. Uotila, L. & Koivusalo, M. in *Coenzymes and Cofactors* (ed. Dolphin, D.) 517–551 (John Wiley & Sons, New York, 1989).
22. Barber, R. D. & Donohue, T. J. Function of a glutathione-dependent formaldehyde dehydrogenase in *Rhodospirillum rubrum* formaldehyde oxidation and assimilation. *Biochemistry* **37**, 530–537 (1998).
23. Ras, J. *et al.* Isolation, sequencing, and mutagenesis of the gene encoding NAD- and glutathione-dependent formaldehyde dehydrogenase (GD-FALDH) from *Paracoccus denitrificans*, in which GD-FALDH is essential for methylotrophic growth. *J. Bacteriol.* **177**, 247–251 (1995).
24. Shen, S., Sulter, G., Jeffries, T. W. & Cregg, J. M. A strong nitrogen source-regulated promoter for controlled expression of foreign genes in the yeast *Pichia pastoris*. *Gene* **216**, 93–102 (1998).
25. Kuwada, M., Horie, S. & Ogura, Y. Studies on the Enzymatic reduction of C-nitroso compounds. II. Multiple forms of liver C-nitrosoreductase and the identity with alcohol dehydrogenase. *J. Biochem. (Tokyo)* **88**, 859–69 (1980).
26. Kluge, I., Gutteck-Amsler, U., Zollinger, M. & Do, K. Q. S-nitrosoglutathione in rat cerebellum: identification and quantification by liquid chromatography-mass spectrometry. *J. Neurochem.* **69**, 2599–2607 (1997).
27. Minning, D. M. *et al.* Ascaris haemoglobin is a nitric oxide-activated 'deoxygenase'. *Nature* **401**, 497–502 (1999).
28. Ribeiro, J. M., Hazzard, J. M., Nussenzveig, R. H., Champagne, D. E. & Walker, F. A. Reversible binding of nitric oxide by a salivary heme protein from a bloodsucking insect. *Science* **260**, 539–541 (1993).
29. Glover, R. E., Koshkin, V., Dunford, H. B. & Mason, R. P. The reaction rates of NO with horseradish peroxidase compounds I and II. *Nitric Oxide* **3**, 439–444 (1999).
30. Liu, L., Zeng, M. & Stamler, J. S. Hemoglobin induction in mouse macrophages. *Proc. Natl Acad. Sci. USA* **96**, 6643–6647 (1999).

Acknowledgements

We thank L. Kane and S. Cutler for help with tetrad dissection and J. Toronto for performing enzyme assays. The work was partly funded by an NIH grant.

Correspondence and requests for materials should be addressed to J.S.S. (e-mail: STAM001@mc.duke.edu).

Covalent inhibition revealed by the crystal structure of the caspase-8/p35 complex

Guozhou Xu^{*}, Maurizio Cirilli^{*†}, Yihua Huang[‡], Rebecca L. Rich[§], David G. Myszka[§] & Hao Wu^{*}

^{*} Department of Biochemistry, Weill Medical College of Cornell University and [†]Program in Physiology, Biophysics and Molecular Medicine, Graduate School of Medical Sciences of Cornell University, 1300 York Avenue, New York, New York 10021, USA

[‡] Istituto di Strutturistica Chimica 'Giordano Giacomello', C.N.R., C. P. 10, Monterotondo Stazione, Italy

[§] Center for Biomolecular Interaction Analysis, School of Medicine, University of Utah, Salt Lake City, Utah 84132, USA

Apoptosis is a highly regulated process that is crucial for normal development and homeostasis of multicellular organisms^{1,2}. The p35 protein from baculoviruses effectively prevents apoptosis by its broad-spectrum caspase inhibition^{3–7}. Here we report the crystal structure of p35 in complex with human caspase-8 at 3.0 Å resolution, and biochemical and mutagenesis studies based on the structural information. The structure reveals that the

caspase is inhibited in the active site through a covalent thioester linkage to p35, which we confirmed by gel electrophoresis, hydroxylamine treatment and mass spectrometry experiments. The p35 protein undergoes dramatic conformational changes on cleavage by the caspase. The repositioning of the amino terminus of p35 into the active site of the caspase eliminates solvent accessibility of the catalytic dyad. This may be crucial for preventing hydrolysis of the thioester intermediate, which is supported by the abrogation of inhibitory activity through mutations at the N terminus of p35. The p35 protein also makes conserved contacts with the caspase outside the active-site region, providing the molecular basis for the broad-spectrum inhibitory activity of this protein. We demonstrate a new molecular mechanism of caspase inhibition, as well as protease inhibition in general.

The p35 protein shows unparalleled effectiveness as a broad-spectrum caspase inhibitor against all three groups of caspases from mammals and other metazoans^{3–6}. Numerous cellular and *in vivo* studies have implicated the function of p35 in rescuing cells from apoptosis to revive their normal cellular functions^{7–12}, thereby

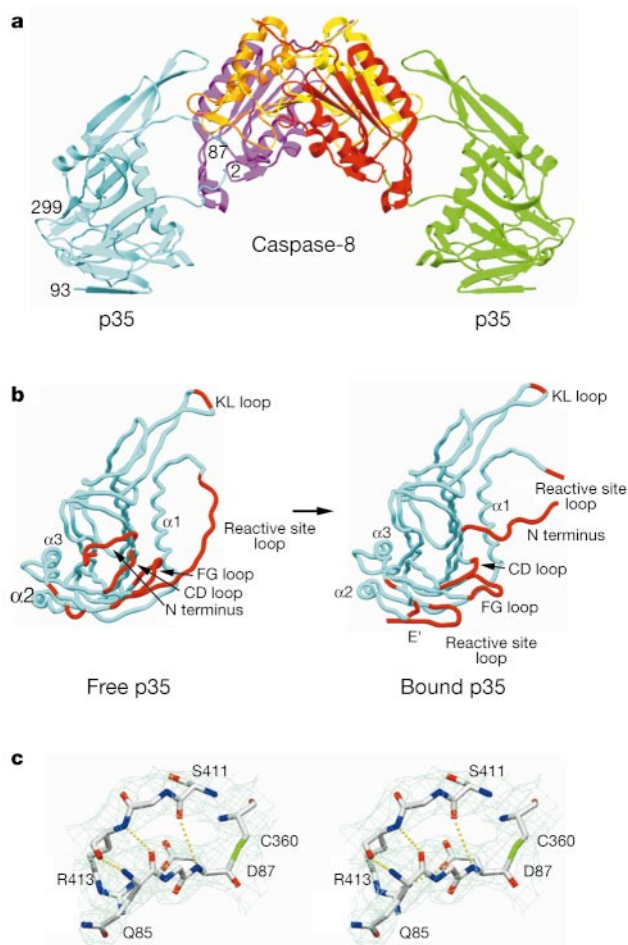


Figure 1 Structure of the p35/caspase-8 complex. **a**, Ribbon diagram of dimeric complex with the two-fold axis in the vertical orientation. p35, cyan and green; α -subunit (p18) of caspase-8, magenta and red; β -subunit (p12) of caspase-8, orange and yellow. Ordered termini for p35-N (residues 2–87) and p35-C (residues 93–299) are labelled.

b, Conformational transitions of p35 on cleavage. Residues with differences in α -carbon positions larger than 4.0 Å are shown in red, which include the N terminus (residues 2–12), the CD loop (residues 35–40), the caspase recognition sequence (residues 85–87), the reactive-site loop after the cleavage site (residues 93–101), the FG loop (residues 157–165) and the KL loop (residues 254–255). **c**, Atomic model of the complex near the active site of caspase-8 overlaid with an omit electron density map (1.0σ contour). Potential hydrogen bonds are indicated by dotted lines. Side chains for residue Met 86 of p35 and Tyr 412 of caspase-8 are omitted for clarity.

establishing the basis for its potential in anti-apoptosis therapies. Biochemical characterization has shown that caspase inhibition by p35 correlates with the cleavage of its reactive-site loop at Asp 87 and the formation of a tight complex with the caspase^{4,6}, by means of an unknown mechanism. To elucidate the molecular basis of p35 function, we determined the crystal structure of the complex between p35 and caspase-8, a group-III-initiating caspase important in apoptosis mediated by tumour necrosis factor (TNF) and Fas (Table 1 and Fig. 1a, b).

Unexpectedly, the electron density map reveals a covalent thioester linkage between the catalytic Cys 360 of caspase-8 and Asp 87 of the bound p35 (Fig. 1c). This is surprising as previous biochemical studies, such as SDS–polyacrylamide gel electrophoresis (SDS–PAGE) experiments, did not observe evidence for a covalent complex between p35 and caspases (refs 6, 13, 14, and data not shown). The crystallographic observation prompted us to re-examine the biochemical experiments. We suspected that the instability of the thioester bond to denaturation treatment might be the reason for the failure to observe this covalent complex on SDS–PAGE. We therefore checked the effects of pH, reducing agent and heat on the integrity of this complex. We found that the complex is destroyed by the heat treatment (100 °C for 3 min) normally recommended for SDS–PAGE¹⁵, irrespective of pH or reducing agent. Under milder heat treatment (80 °C for 2 min), however, the complex exhibits a distinct dependence on reducing agent and pH (Fig. 2a), which is consistent with the stability and chemistry of a thioester bond

Table 1 Refinement statistics

Resolution	40–3.0 Å
R_{sym} : overall (last shell)	9.5% (28.9%)
Coverage: overall (last shell)	94.9% (81.7%)
R (free R)	23.6% (29.6%)
Number of reflections used	36,808
Number of protein residues	1,070
Number of protein atoms	8,696
Number of solvent atoms	26

$$R_{\text{sym}} = \sum_i \sum_j |I_{ij} - \langle I_{ij} \rangle| / \sum_i \sum_j I_{ij}$$

between p35 and the caspase¹⁶. Additional experimental corroboration of the presence of the covalent thioester bond was obtained from the sensitivity of the complex to treatment by hydroxylamine (Fig. 2b), a strong nucleophile often used to break ester bonds^{16,17}.

The covalent nature of the interaction between p35 and caspases was further confirmed by matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry (Fig. 2c), which rarely preserves non-covalent interactions. The observed spectrum of the p35/caspase-8 complex contains a prominent peak at around a relative molecular mass of 28,000 (M_r 28K) that can only be explained by a covalently linked N-terminal fragment of p35 (p35-N, residues 2–87) and the α -subunit of caspase-8 (p18). Consistent with this interpretation, observed peaks for the two individual components (p35-N and p18) are very weak. Furthermore, acid denaturation of the complex in the presence of reducing agent eliminated the 28K peak and enhanced the peaks for the individual components (p35-N and p18). The MALDI-TOF spectrum of the complex of p35 with caspase-3, a group-II-executing caspase, contains a similar peak of high molecular mass, which suggests the generality of the covalent interaction of p35 with caspases.

The trapping of the thioester intermediate between p35 and caspases raises the question as to how the intermediate is stabilized rather than hydrolysed by the enzyme. The structure of the complex suggests that the N terminus of p35 may have a specific and crucial function in this regard by direct blockade of hydrolysis. During normal catalysis, the thioester intermediate is quickly hydrolysed by an activated water molecule bound to the His residue in the catalytic dyad¹⁸. In the complex, the N terminus of p35 repositions into the active site of caspase-8 from a partially buried conformation in the non-cleaved p35 (Figs 1b and 3a). The interaction eliminates completely solvent accessibility of the critical His 317 residue in caspase-8, which is semi-exposed in the absence of the interaction. Although the resolution of the structure does not permit a definitive assignment, the thiol of Cys 2 also seems to form a hydrogen bond with the imidazole ring of His 317, which distorts His 317 away from an optimal orientation for catalysis. Residue Cys 2—which we demonstrated by mass spectrometry of tryptic peptides to be capped by acetylation—and residue Val 3 make van der Waals contacts with caspase-8, at residues Ile 257, Asp 363 and Tyr 365 (Fig. 3a). The interaction does not seem to affect otherwise the intrinsic catalytic machinery of caspase-8, as the p35/caspase-8 structure superimposes well with other known caspase-8 structures^{19,20}.

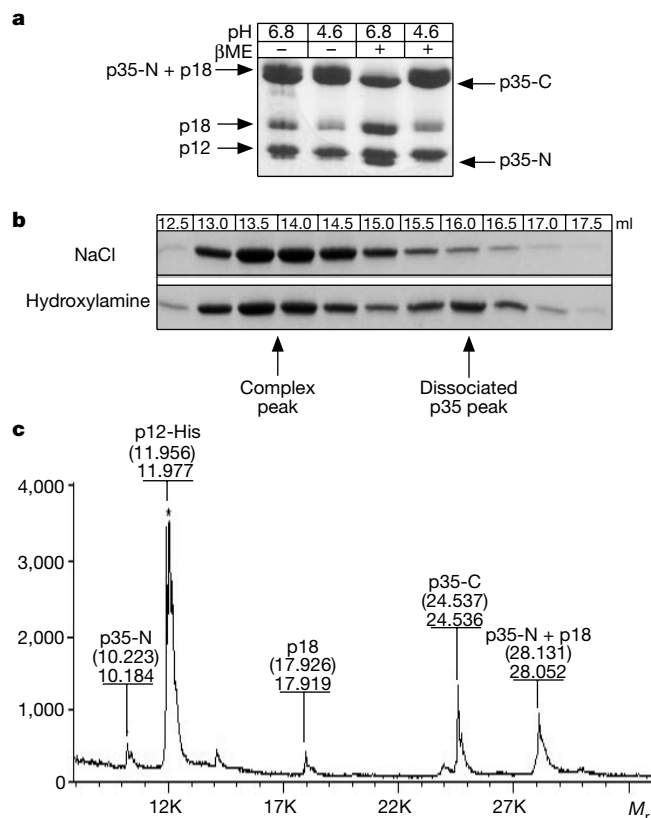


Figure 2 Biochemical characterization of the p35/caspase-8 complex. **a**, SDS–PAGE of the complex treated with SDS sample buffers at different pH levels and with or without the reducing agent β -mercaptoethanol (β ME). **b**, Gel-filtration profiles of the complexes, treated either with 1.2 M hydroxylamine or 1.2 M NaCl for 2 h at 37 °C, showing the dissociation of p35 on hydroxylamine treatment. For clarity, only the bands for p35–C are shown. **c**, Observed MALDI–TOF mass spectrometry spectrum of the complex. P35–N, residues 2–87; p35–C, residues 88–299. The observed and calculated (in parentheses) molecular masses for each peak are shown. The values on the y axis are arbitrary units. Asterisk, truncated peak.

Table 2 Characterization of the interaction between p35 and caspases

p35	Complex formation*	Cleavage*	Interaction with caspase-3 (C163A)†
Wild type	+	+	$K_d = 1.16 \pm 0.01 \times 10^{-7}$ M
C2G	–	+	
V3G	–	+	
Δ 2–5	–	+	
Y82A	–	+	$K_d = 3.4 \pm 0.1 \times 10^{-5}$ M

K_d , dissociation constant.

* Complex formation and cleavage by both caspase-3 and caspase-8 were determined by pull-down of caspases using His-tagged p35 proteins.

† Determined by biosensor measurement (see Supplementary Information).

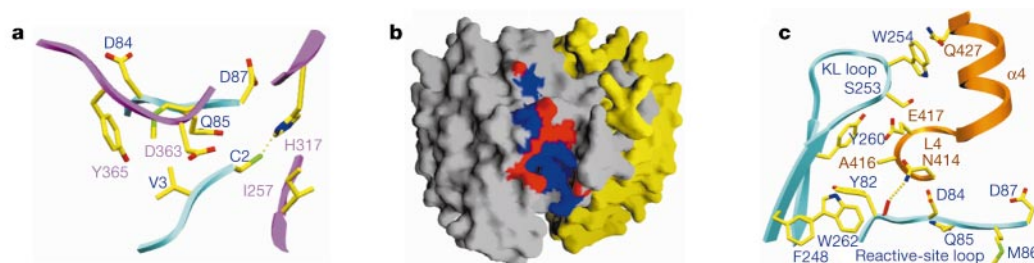


Figure 3 Detailed interaction between p35 and caspase-8. **a**, Interaction near the N terminus of p35 with the active site of caspase-8. p35, cyan; α -subunit of caspase-8, magenta. **b**, Footprint of p35 on the surface of caspase-8 (grey and yellow for each of the

$\alpha\beta$ -units). Residues conserved among different caspases and main-chain atoms are shown in blue, whereas those with significant variations are shown in red. **c**, Interaction near the KL loop of p35. p35, cyan; β -subunit of caspase-8, orange.

The importance of the N terminus of p35 in the inhibitory mechanism was confirmed by mutagenesis experiments. We created two point mutants (C2G and V3G) and a deletion mutant ($\Delta 2-5$) at the N terminus. None of the mutants could form stable complexes with either caspase-8 or caspase-3, based on His-tag pull-down experiments (Table 2). However, all three mutants can be cleaved by both caspases. The C2G mutant is an efficient caspase substrate and exhibits identical behaviour to that of wild-type p35 in expression, solubility and gel-filtration profile. Therefore, the mutational effect of C2G can be attributed to a direct deletion of the important interaction in the complex rather than through significant perturbation of the structure of p35. The V3G and $\Delta 2-5$ mutants were cleaved efficiently by caspase-3, but less efficiently by caspase-8. Our gel-filtration experiments showed that they are predominantly dimeric in solution, which is in contrast to monomers for wild-type p35 and the C2G mutant.

Structural analysis of the p35/caspase-8 complex shows that the binding of the p35 N terminus into the active site of the caspase is made possible by the cascade of conformational changes in p35 on its cleavage (Fig. 1b). In the free p35 structure, there is a sequential packing interaction from the reactive-site loop to the FG loop, the CD loop and the N-terminal segment. The pinching of the reactive-site loop in the bound p35 structure suggests that this loop may be highly strained in the initial encounter with the caspase. Therefore, residues after the cleavage site may immediately spring away from the caspase on cleavage. This should instantly cause the sequential relaxation of the FG loop and the CD loop and the release of the N terminus from the core p35 structure to reposition into the active site of the caspase. Side chains adjacent to the N terminus, such as those in helices $\alpha 2$ and $\alpha 3$, move in to fill part of the vacated space. The portion of the reactive-site loop after the cleavage site folds into an extra β -strand (denoted E') next to the E strand at the edge of the p35 β -sandwich.

These structural transitions in caspase inhibition by p35 explain previous kinetic measurements on the characteristic slow-binding inhibition of p35 (ref. 6) in which the initial encounter is followed presumably by a slower isomerization step. The time scale of the conformational changes must be such that thioester hydrolysis does not occur in the interim between the departure of the P' residues and the binding of the N terminus into the active site. These two processes are probably highly cooperative, as the strain in the reactive-site loop may pull the P' residues away as soon as cleavage occurs to allow repositioning of the N terminus into the active site. However, there may be some degree of leakage in the mechanism. This is supported by the apparent deviation from a 1:1 stoichiometry of p35 inhibition for some caspases and the possible turn-overs of p35 by a substrate pathway⁶.

The non-covalent interaction between p35 and caspase-8 centres on and extends beyond the tetrapeptide caspase recognition sequence (P4-D⁸⁴QMD⁸⁷-P1) in the reactive-site loop. The contact surface shows significant conservation among different caspases

(Fig. 3b), providing an explanation for the wide-spectrum effectiveness of p35 as a caspase inhibitor. The total interaction buries approximately 950 Å² of surface area on each partner of the binding interface. The P4-P1 residues are recognized similarly by the S4-S1 sites, as seen in crystal structures of caspases in complex with various peptide inhibitors^{19,20}, involving specific main-chain and side-chain hydrogen bonds with conserved caspase residues such as Arg 260, Ser 411 and Arg 413. The N terminus of p35 contacts the caspase adjacent to one side of this interaction (Fig. 3a). To the other side, the P6 residue Tyr 82, residues in the K and L strands, and the KL loop interact with residues 414-427 of caspase-8 at the structurally conserved L4 loop and the $\alpha 4$ helix (Fig. 3c). These caspase residues show a lesser degree of conservation at the sequence level, but significant main-chain contacts are used. It is probable that the observed flexibility in the KL loop and the possible flexibility of the N terminus of p35 can accommodate some degree of sequence variation to preserve the interaction.

As the mechanism of caspase inhibition by p35 involves integration of multiple sequence segments, the integrity of the p35 structure is crucial for its function as a caspase inhibitor. Our structure-based alanine mutagenesis for p35 interface residues showed that Tyr 82 at the base of the reactive-site loop is essential for complex formation with caspase-3 and caspase-8, without affecting cleavage. The strength of the initial encountering interaction is also greatly diminished, as assessed by the biosensor measurement of the interaction between p35 and an active site mutant of caspase-3 (C163A), which was processed to its mature form by caspase-8 (Table 2). Tyr 82 helps to glue together the reactive-site loop and the neighbouring KL strands and loop. Its mutation may therefore uncouple these interaction elements for caspases. Disruptive mutations at residues such as Ile 67 and Val 71 have shown similar phenotypes and seem to act by affecting the interaction between the β -sheet core of p35 and the helix ($\alpha 1$) preceding the reactive-site loop¹⁴. Many reported Ala-Ser insertions and alanine mutagenesis of charged residues in p35 may also be explained by structural perturbations^{14,21}.

The molecular mechanism of caspase inhibition by p35 reported here is highly distinct from the collection of modes of protease inhibition observed in numerous protease-inhibitor complexes²². Most of these protease inhibitors block the active sites of their target proteases either directly or indirectly through extensive non-covalent interactions. Covalent suicide inhibition through ester-bond formation has long been known for serpins, a family of serine protease inhibitors²³, but with a different underlying mechanism. The recent crystal structure of a serpin/protease complex shows distortion of the active site and deformation of the protease by a pulling force from the serpin²⁴. In contrast, p35 directly blocks thioester hydrolysis after post-cleavage conformational changes through solvent exclusion and perhaps reorientation of the catalytic residue, providing yet another new mode of protease inhibition. □

Methods

Protein expression and purification

Baculovirus p35 (residues 1–299, with and without a C-terminal His-tag), human caspase-8 (residues S201–D463, with and without a C-terminal His-tag), human caspase-3 (residues 1–277, with and without a C-terminal His-tag) and human caspase-3 active site mutant (C163A; residues 1–277, with a C-terminal His-tag) were expressed in the pET bacterial expression system using isopropylthiogalactoside induction overnight at 20 °C. The His-tagged proteins were purified by Ni-affinity chromatography and gel filtration. Both caspase-8 and caspase-3 were automatically converted to their mature forms during the protein expression and purification procedures. To obtain the p35/caspase-8 complex, the cell pellets of His-tagged caspase-8 were mixed with those containing excess non-tagged p35, and the complex was co-purified using the His-tag in the caspase-8 construct. To generate the mature form of the mutant caspase-3 (C163A), the cell pellets of His-tagged C163A mutant were mixed with those of non-tagged caspase-8. The mixed cells were lysed by repeated bursts of sonication and incubated for 30 min at 37 °C to allow the processing of the C163A mutant by the active caspase-8 in the lysate. The processed C163A mutant was then purified similarly by Ni-affinity and gel filtration.

Crystallization of the p35/caspase-8 complex

The p35/caspase-8 complex eluted from a Superdex 200 gel-filtration column at a position around 130K, corresponding to a complex containing two molecules of p35 and one dimeric $\alpha_2\beta_2$ caspase-8 (α -subunit-p18 and β -subunit-p12). The gel-filtration buffer contained 20 mM Tris at pH 7.5, 150 mM NaCl and 5 mM dithiothreitol. The protein complex was concentrated to 20–30 mg ml⁻¹ and crystallized at 20 °C by hanging-drop vapour diffusion. The crystallization reservoir contained 0–2% PEG8K, 100 mM MOPS at pH 7.5 and 1% 2-propanol. The reduction in salt concentration in the crystallization drop after mixing the complex with the reservoir seems to have been the principal force that decreased the solubility of the complex to allow saturation and crystallization.

Data collection and structural determination

The crystals belong to space group C22₁ with cell dimensions $a = 100.0$ Å, $b = 117.3$ Å and $c = 346.5$ Å. Although the crystals looked indistinguishable, more than 95% of the crystals are intrinsically twinned. An initial 3.5-Å diffraction data set from a single crystal was collected at the X4A beamline of National Synchrotron Light Source, from which a molecular replacement solution was obtained using the atomic models of free p35 (ref. 13) and the $\alpha\beta$ -form of a caspase-8 structure²⁰ using the program Replace²⁵. Each asymmetric unit of the crystals contains two p35 molecules and two separate $\alpha\beta$ -units of caspase-8. The two-fold axes of the caspase-8 dimers coincide with the crystallographic axes. A 3.0-Å data set was collected at the Commercial Collaborative Access Team of the advanced photon source for model building²⁶ and refinement²⁷. Marked conformational rearrangements were observed for p35 and built on the basis of $2F_o - F_c$ maps and averaged electron density maps.

Mass spectrometry

MALDI-TOF analyses were performed in the linear positive-ion mode with delayed ion extraction using a Biflex III mass spectrometer (Bruker Daltonik). Desalted samples in 0.1% aqueous trifluoroacetic acid (TFA) were mixed 1/1 (v/v) with a matrix consisting of 10 mg ml⁻¹ 2,5-dihydroxybenzoic acid dissolved in acetonitrile/water/TFA (50/50/0.1 (v/v/v)) and spotted on a highly polished stainless target. Spectra were calibrated using the singly and doubly charged peaks of horse skeletal muscle apomyoglobin as either internal or external standards. We collected and analysed spectra using proprietary Bruker software.

Biosensor measurement

The interaction of p35 with the mature form of the active-site mutant of caspase-3 (C163A), which was immobilized on the sensor chip by amine-coupling chemistry, was measured by surface plasmon resonance using a BIACORE 2000. The experiment was performed similarly as described for the interaction between TRAF2 and TRADD²⁸. All response data were double referenced²⁹, thereby correcting for bulk refractive index changes and nonspecific p35 binding to the sample and reference surfaces. Data were fitted globally to a simple interaction model ($A + B = AB$) using CLAMP³⁰.

His-tag pull-down assay

Purified His-tagged wild-type or mutant p35 proteins were mixed with cell pellets of non-tagged caspase-8 or caspase-3 and lysed in 50 mM phosphate at pH 8.0, 150 mM NaCl, 10 mM imidazole and 10 mM β -mercaptoethanol. Pre-equilibrated Ni-NTA resins were incubated with the mixtures at 4 °C overnight and washed three times with the same buffer. The bound proteins were eluted with 200 mM imidazole and analysed on SDS-PAGE.

Received 16 January; accepted 26 February 2001.

1. Thompson, C. B. Apoptosis in the pathogenesis and treatment of disease. *Science* **267**, 1456–1461 (1995).
2. Steller, H. Mechanisms and genes of cellular suicide. *Science* **267**, 1445–1449 (1995).
3. Clem, R. J., Fechheimer, M. & Miller, L. K. Prevention of apoptosis by a baculovirus gene during infection of insect cells. *Science* **254**, 1388–1390 (1991).
4. Bump, N. J. *et al.* Inhibition of ICE family proteases by baculovirus antiapoptotic protein p35. *Science* **269**, 1885–1888 (1995).
5. Xue, D. & Horvitz, H. R. Inhibition of the *Caenorhabditis elegans* cell-death protease CED-3 by a CED-3 cleavage site in baculovirus p35 protein. *Nature* **377**, 248–251 (1995).
6. Zhou, Q. *et al.* Interaction of the baculovirus anti-apoptotic protein p35 with caspases. Specificity, kinetics, and characterization of the caspase/p35 complex. *Biochemistry* **37**, 10757–10765 (1998).
7. Ekert, P. G., Silke, J. & Vaux, D. L. Caspase inhibitors. *Cell Death Differ.* **6**, 1081–1086 (1999).
8. Hay, B. A., Wolff, T. & Rubin, G. M. Expression of baculovirus P35 prevents cell death in *Drosophila*. *Development* **120**, 2121–2129 (1994).
9. White, K., Tahaoglu, E. & Steller, H. Cell killing by the *Drosophila* gene reaper. *Science* **271**, 805–807 (1996).
10. Beidler, D. R., Tewari, M., Friesen, P. D., Poirier, G. & Dixit, V. M. The baculovirus p35 protein inhibits Fas- and tumor necrosis factor-induced apoptosis. *J. Biol. Chem.* **270**, 16526–16528 (1995).
11. Robertson, N. M. *et al.* Baculovirus P35 inhibits the glucocorticoid-mediated pathway of cell death. *Cancer Res.* **57**, 43–47 (1997).
12. Hisahara, S. *et al.* Targeted expression of baculovirus p35 caspase inhibitor in oligodendrocytes protects mice against autoimmune-mediated demyelination. *EMBO J.* **19**, 341–348 (2000).
13. Fisher, A. J., Cruz, W., Zoog, S. J., Schneider, C. L. & Friesen, P. D. Crystal structure of baculovirus P35: role of a novel reactive site loop in apoptotic caspase inhibition. *EMBO J.* **18**, 2031–2039 (1999).
14. Zoog, S. J., Bertin, J. & Friesen, P. D. Caspase inhibition by baculovirus P35 requires interaction between the reactive site loop and the beta-sheet core. *J. Biol. Chem.* **274**, 25995–26002 (1999).
15. Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular Cloning*, 2nd edn (Cold Spring Harbor Laboratory Press, New York, 1989).
16. Bruice, T. C. & Benkovic, S. J. *Bioorganic Mechanisms* (Benjamin, New York, 1966).
17. Owen, W. G., Penick, G. D., Yoder, E. & Poole, B. L. Evidence for an ester bond between thrombin and heparin cofactor. *Thromb. Haemost.* **35**, 87–95 (1976).
18. Stennicke, H. R. & Salvesen, G. S. Catalytic properties of the caspases. *Cell Death Differ.* **6**, 1054–1059 (1999).
19. Watt, W. *et al.* The atomic-resolution structure of human caspase-8, a key activator of apoptosis. *Structure Fold Des.* **7**, 1135–1143 (1999).
20. Blanchard, H. *et al.* The three-dimensional structure of caspase-8: an initiator enzyme in apoptosis. *Structure Fold Des.* **7**, 1125–1133 (1999).
21. Bertin, J. *et al.* Apoptotic suppression by baculovirus P35 involves cleavage by and inhibition of a virus-induced CED-3/ICE-like protease. *J. Virol.* **70**, 6251–6259 (1996).
22. Bode, W. & Huber, R. Structural basis of the endoprotease–protein inhibitor interaction. *Biochim. Biophys. Acta* **1477**, 241–252 (2000).
23. Wright, H. T. & Scarsdale, J. N. Structural basis for serpin inhibitor activity. *Proteins* **22**, 210–225 (1995).
24. Huntington, J. A., Read, R. J. & Carrell, R. W. Structure of a serpin–protease complex shows inhibition by deformation. *Nature* **407**, 923–926 (2000).
25. Tong, L. REPLACE, a suite of computer programs for molecular-replacement calculations. *J. Appl. Cryst.* **26**, 748–751 (1993).
26. Jones, T. A., Zou, J. -Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building models in electron density maps and the location of errors in those models. *Acta Crystallogr. A* **47**, 110–119 (1991).
27. Brunger, A. T. *et al.* Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
28. Park, Y. C. *et al.* A novel mechanism of TRAF signaling revealed by structural and functional analyses of the TRADD–TRAF2 interaction. *Cell* **101**, 777–787 (2000).
29. Myszk, D. G. Improving biosensor analysis. *Mol. Recogn.* **12**, 1–6 (1999).
30. Myszk, D. G. & Morton, T. A. CLAMP: a biosensor kinetic data analysis program. *Trends Biochem. Sci.* **23**, 149–150 (1998).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank K. D'Amico for access to the COM-CAT beamline at the APS; C. Ogata for access to the X4A beamline at NSLS; L. Tong for help with data collection and for critical reading of the manuscript; J. Kerwin for mass spectrometry experiments; J. Luft and G. DeTitta for initial screening of crystallization conditions for a p35/caspase-3 complex, which served as leads for the crystallization of the p35/caspase-8 complex; and C. Lima, T. Muir, B. Chait, G. Dodson, H. T. Wright and T. Begley for insightful discussions. This work was supported by the Speaker's Fund for Biomedical Research and the departmental startup fund. H.W. is a Pew Scholar in the Biomedical Sciences.

Correspondence and requests for materials should be addressed to H.W. (e-mail: haowu@med.cornell.edu). Atomic Coordinates are deposited in the Protein DataBank under accession number 1I4E.